US/9646GO

us-09-340-736-1.rag

```
                    GenCore version 4.5
            Copyright (c) 1993 -(2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      April 24, 2001, 16:31:20 ; Search time 115.25 seconds
                                        (without alignments)
                                        362.571 Million cell updates/sec

Title:       US-09-340-736-1
Perfect score: 3785
Sequence:    1 GCVPGAIPGGVPGGVFYPG2..........LSPIFPGGACLGKACGRXRK 731

Scoring table: BLOSUM62
            GapOp 10.0 , GapExt 0.5

ched:       390729 seqs, 5716235 residues

Total number of hits satisfying chosen parameters:          390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries
```

Database :   A_Geneseq_0401:*
```
 1- /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
 2- /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
 3- /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
 4- /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
 5- /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
 6- /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
 7- /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
 8- /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
 9- /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10- /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11- /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12- /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13- /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14- /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15- /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16- /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17- /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18- /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19- /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20- /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21- /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22- /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3785	100.0	731	21	Y69068	Amino acid sequenc
2	3785	100.0	733	15	R56653	Synthetic human tr
3	3785	100.0	733	20	Y01301	Amino acid sequenc
4	3768.5	99.6	730	19	W46315	Human elastin cont
5	3588.5	94.8	698	20	Y01302	Human tropoelastin
6	3588.5	94.8	698	21	Y69069	Amino acid sequenc
7	3458.5	91.4	712	21	B08630	Amino acid sequenc
8	3458.5	91.4	730	21	B08631	Fusion protein com
9	3222	85.1	660	20	Y01303	Human tropoelastin
10	2869	75.8	571	21	Y69071	Amino acid sequenc
11	2680	70.8	515	21	Y69135	Amino acid sequenc

12	1159.5	30.6	988	16	R80253	Polymer SELP7. Sy
13	1159	30.6	832	16	R80252	Polymer SELP8. Sy
14	1130	29.9	936	16	R80251	Polymer SELP0. Sy
15	1126.5	29.8	877	16	R80335	protein polymeric
16	1126.5	29.8	877	16	W49724	protein polymer ad
17	1123	29.7	884	16	R80341	protein polymer ad
18	1123	29.7	884	17	W09213	SELP8K polymer. S
19	1123	29.7	884	19	W49728	SELP8K polymer. S
20	1123	29.7	884	19	W53541	Expected amino aci
21	1123	29.6	884	21	Y51882	plasmid pPT0345 pr
22	1122	29.6	1002	17	W09218	SELP0K polymer. S
23	1122	29.6	1002	19	W53544	Amino acid sequenc
24	1122	29.6	1002	21	Y51888	plasmid pPT0364 pr
25	1118	29.5	1412	19	W53519	Amino acid sequenc
26	1118	29.5	1413	9	P82957	ESBI protein compr
27	1118	29.5	1413	14	R41008	ESBI multimeric pr
28	1118	29.5	1413	18	W26343	ESBI synthetic ela
29	1118	29.5	1464	21	Y78278	ESBI protein seque
30	1116.5	29.5	745	20	Y31682	Alanine-containing
31	1116	29.5	1056	16	R80254	Polymer SELP3. Sy
32	1107	29.2	2257	9	P82961	SEPI3 protein comp
33	1107	29.2	2257	14	R41012	SELP3 multimeric p
34	1107	29.2	2257	18	W26347	SELP3 synthetic el
35	1107	29.2	2257	19	W53523	Amino acid sequenc
36	1107	29.2	2257	21	Y78282	SELP3 amino acid s
37	1105	29.2	216	20	Y01310	Human tropoelastin
38	1100	29.1	936	17	W09221	SELP0K-CSI polymer
39	1100	29.1	936	21	Y51891	plasmid pPT0370 pr
40	1100	29.1	937	19	W53547	Amino acid sequenc
41	1095	28.9	972	16	R80255	Polymer SELP4. Sy
42	1070.5	28.3	2055	9	P82960	SEPI2 protein comp
43	1070.5	28.3	2055	14	R41011	SELP2 multimeric p
44	1070.5	28.3	2055	18	W26346	SELP2 synthetic el
45	1070.5	28.3	2055	19	W53522	Amino acid sequenc

ALIGNMENTS

```
RESULT    1
Y69068
ID   Y69068 standard; Protein; 731 AA.
XX
AC   Y69068;
XX
DT   30-MAY-2000  (first entry)
XX
DE   Amino acid sequence of a human tropoelastin splice form.
XX
KW   Tropoelastin; derivative; proteolysis; protease; antiwrinkle;
KW   hand lotion; bulking agent; chemotaxis; proliferation;
KW   growth inhibition; peptidomimetic; lung damage; elastin; cancer;
KW   metastasis; blood clotting.
XX
OS   Homo sapiens.
XX
FH                   Key              Location/Qualifiers
FT   Cleavage-site                   515..516
FT   Cleavage-site                   564..565
FT   Cleavage-site                   441..442
FT   Cleavage-site                   503..504
XX
PN   WO20004043-A1.
XX
PD   27-JAN-2000.
XX
PF   19-JUL-1999;     99WO-AU00580.
XX
PR   17-JUL-1998;     98AU-0004723.
XX
PA   (UNSI ) UNIV SYDNEY.
XX
PI   Weiss AS;
```

```
XX
DR   WPI; 2000-182399/16.
DR   N-PSDB; Z61146.
XX
PT   New derivatives of tropoelastin useful, e.g. for inducing chemotaxis
PT   and cell growth -
XX
PS   Disclosure; Page 107-109; 135pp; English.
XX
CC   The present sequence represents a human tropoelastin splice form. The
CC   specification describes tropoelastin derivatives, in which a subsequence
CC   has been mutated so that susceptibility to proteolysis is reduced or
CC   eliminated, or a subsequence has been inserted so that susceptibility
CC   to proteolysis is increased. The derivatives have with reduced
CC   susceptibility, and can be used where the wild-type protein would be
CC   degraded too easily, e.g. in contact with serum or wound exudate. The
CC   tropoelastin derivatives provide competitive inhibition of protease
CC   activity. The tropoelastin derivatives, and other polypeptides containing
CC   tropoelastin derivative-derived protease-susceptibility sites, are useful
CC   in human or veterinary medicine, cosmetics (e.g. antiwrinkle or hand
CC   lotions), as bulking agents and for inducing chemotaxis. They are also
CC   useful for proliferation or growth inhibition, particularly of smooth
CC   muscle cells, epithelial or endothelial cells, fibroblasts, osteocytes,
CC   chondrocytes and platelets. Peptidomimetics that mimic the protease
CC   cleavage site in tropoelastin derivatives are competitive inhibitors of
CC   the protease, and are used for protecting against lung damage caused by
CC   elastin, for inhibiting or controlling localised growth of cancers or
CC   metastases, or to limit protease activity that causes blood clotting.
XX
SQ   Sequence   731 AA;
```

```
Query Match             100.0%;  Score 3785;  DB 21;  Length 731;
Best Local Similarity   100.0%;  Pred. No. 3e-213;
Matches  731;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1  GSVPGAIPGGVPGCGVTYPGAGLGALGGALGPGGKPLKPVPGGLAGASLGAGLGAFPAVT   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  ggvpgaipggvpggvfypgaglgalgggalgpggkplkpvpgglagaglgaglgafpavt   60

Qy     61  FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVPQPGAGVKPGKVPGVGL  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  fpgalvpggvadaaaaykaakagaglggvpgvgglgvsagavpqpgagvkpgkvpgvgl  120

Qy    121  PGVYPGGVLPGARFPGVGVLPGVFIGAGYXPKAPGVGGAFAGIPGVGPFGGPGPGVPLGY  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  pgvypggvlpgarfpgvgvlpgvgvgtgagvXpkapgvggafagipgvgpfggpgpgvpgy  180

Qy    181  PIKAPKLPGGYGLPYTTGKLPYTGYPGGYAGRAGKAGPFTGYGVGPGPCAAAAAAKAAAKF  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  pikapklpggygylpyttgklpytgypgggyagraagkagyfptgtgyvgpgaaaaaaakf  240

Qy    241  GAGAGVLPGVGAGVPGVPGGAIPGIGSGIAGVGPAAAAAAAAAKAKYGAAGLVPGG  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  gagaagvlpgvgagvpgvpggaipgigsgiagvgpaaaaaaaaakaakygaaagivpgg  300

Qy    301  PGFGPGVVGPGAGVPGVGVPGVGVGGFPGFGVGVGGIPGVAGVPGVSGVGGVPGVGGIPGVGGVGVGGGPGVPLGT  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  pgfgpgvvgpgagvpgvgvpgvgvgggfpgfgvgvgggipgvsgvgggpgvgvgggpgvggar  360

Qy    361  PGVGVGGIPTYGVGAGGFPGFGVGVGGIPGVAGVPGVGVPGVSPEAQAAAA  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  pgvgvgglptygvgaggfpgfgvgvgglpgvagvpgvsgvgvpgvispeaqaaaa  420

Qy    421  AKAAKYGVGGIPAAAAAKAAAQFGLVPGVGVARPGVAPGVGVAPGVGVAPG  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  akaakygvgtpaaaaaakaaaaqfglvpgvgvapgvgvapgvgvapgvgap  480

Qy    481  VGVAPGVGVAPGIGPGGVVAAAAKSAAKYAAKAQLRAAAGLAGAPGLGLGVGVPGLGVGA  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  vgtapgvgvapglgpggvaaaaaaksaakyaakaqlraaaglgagglpglgvgvpglgvga  540

Qy    541  GVPGLGVGAGVPGFGAGADGGVRRSLSPELIREGDPSSQHLPSTPSSFRVPCALAAAKA  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    541  gvpglgvgagvpgfgagdegvrrslspelrregdpssqhlpstpssprvpgalaaakaa  600

Qy    601  KYGAAVPGVLGLGALGGVGIPGGVVGAGPAAAAAAKAAKAAQPGLVGAAGLGGLGVG  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  kygaavpgvlgglgalggvgipggvvgagpaaaaaaakaaakaaqpglvgaaglgglgvg  660

Qy    661  GLGVPGVGGLGGIPPAAAAKAAKYGAAGLGGVLGGANGQPFLGGVAARPGFGLSPIFPGGA  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  glgvpgvggylgglppaaaakaaakygaaaglggvlggagiggvlggagfpiggvaarpgfglspifpgga  720

Qy    721  CLGKACGRKKK 731
           |||||||||||
Db    721  clgkacgrkrk 731
```

```
RESULT   2
R56653
ID   R56653 standard; Protein; 733 AA.
XX
AC   R56653;
XX
DT   22-MAR-1995 (first entry)
XX
DE   Synthetic human tropoelastin (SHEL).
XX
KW   Tropoelastin; pharmaceutical; surgical dressing.
XX
OS   Synthetic.
XX
PN   WO9414958-A.
XX
PD   07-JUL-1994.
XX
PF   16-DEC-1993;  93WO-AO00655.
XX
PR   22-DEC-1992;  92AU-0005520.
PR   28-JUN-1993;  93AU-0009661.
XX
PA   (UNSI ) UNIV SYDNEY.
XX
PI   Martin SL, Weiss AS;
XX
DR   WPI; 1994-263633/32.
DR   N-PSDB; Q70941.
XX
PT   Synthetic polynucleotide(s) - encode recombinant tropoelastins
PT   and variants
XX
PS   Disclosure; Page 30; 77pp; English.
XX
CC   Human synthetic tropoelastin is susceptible to hydrolytic breakdown
CC   of the crosslinks. Such material may be useful in e.g. surgical
CC   applications, where the gradual loss of material over time is
CC   intended.
XX
SQ   Sequence   733 AA;
```

```
Query Match             100.0%;  Score 3785;  DB 15;  Length 733;
Best Local Similarity   100.0%;  Pred. No. 3e-213;
Matches  731;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1  GSVPGAIPGGVPGCGVTYPGAGLGALGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3  ggvpgaipggvpggvfypgaglgaglgggalgpggkplkpvpgglagaglgaglgafpavt   62

Qy     61  FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVPQPGAGVKPGKVPGVGL  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63  fpgalvpggvadaaaaykaakagagglgggvpgvgggvgvsagavvpqpgqpgagvkpgkvpgvgl  122

Qy    121  PGVYPGGVLPGARFPGVGVLPGVFIGAGYXPKAPGVGGAFAGIPGVGPFGGPGPGVPLGY  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   123  psYvpYpggYvlpgarifpgvyvlpgYvpgagvYpkapgvygafagipvgvpfSggpapgvYplgy  182

Qy   181  PIKAPKLPGGYGLPYTTGKLPYGYGPGGVYAGRAGKAGYPTGTGVGPQAAAAAAAAKAAAKF       240
               ||||||||||||  |||| ||  | || || ||||||  |||||||||||||||||
Db   183  pikapklpggygylpyttgklpyvyvgypgygyvpgaaaagkagypgtgtgvgpgaaaaaaakaaakf   242

Qy   241  GAGRAGYLPGVGGACVPGGAIPGIGGIAGVGTPAAAAAAAAAAAKYGAAAGLVPGG              300
               |||||| || ||   ||||| |||| |||| ||| ||||||||||||||||||||
Db   243  gagaagvlpgvyggvfypgagipgaigglagvvgtpaaaaaaaaaakaaKygaaagivpgg         302

Qy   301  PGFGPGVGVVPGAGVPGVGVGPGAGIPVVPGAGIPGAAVPGVVSPEAAAAKAAAAKYGAR          360
               |||||||||  ||||||||| |||||||  ||||||||||||||||||||||||||
Db   303  pgfgpgvvvvpgagvpgvgvgpgagipvvpgagvpgaavpgvvspeaaakaaakaaKygar         362

Qy   361  PGVGVGGIPYTVGVGAGGFPGFGVGVGGIPVAGVPSVGGVPGVGVGVGVGVGISPEAQAAAA        420
               |||||||||||  |||| |||||||||||||||||||||||||||||||||||||||||
Db   363  pgvgvggipYtvgvgaggfpgfgvgvggipvagvpsvggvpgvgvgvgvgvgvgispeaqaaaa     422

Qy   421  AKAAKYGVGTPAAAAAAKAAKAAQFGLVPGVGVGVGVAPGVGLAPGVGVAPG                  480
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   423  akaakygvgtpaaaaaakaaakaaQfgivpgvgvgvgvapgvgvapgvglapgvgvapg           482

Qy   481  VGVAPGVGVAPGIGPGGVAAAAKSAAKVTAKAQLRAAAGLGAGIPGLGVGVGPGLGVGA          540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   483  vgvapgvgvapgigpggvaaaaaksaakvaakaglraaaglgagipglgvgvgvpglgvga         542

Qy   541  GVPGLGVGAGVPGPGAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAA         600
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   543  gvpglgvgagvpgfgagadegvrrslspelregdpssqhlpstpsSprvpgalaaakaa          602

Qy   601  KYGAAVPGVLGGLGALGGVGIPGGVGGIPGGVPAAAAKAAKYGAAGLGVLGAGQFPLGGVAARPFGLSPIFPGGA   720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   603  KygaavpgvlggIgalggvgipggvggipggvpaaaaKaaKyga                          662

Qy   661  GLGYPGVGGLGIPPAAAAKAAKYGAAGLGGVLGAGQFPLGGVAARPFGLSPIFPGGA             720
               ||||||||||||||||||
Db   663  glgvpgvggIgippaaaaKaaKygaagfgivgalggvgipggvvggipggga                  722

Qy   721  CLGKACGRKRR 731
               ||||||||||
Db   723  clgkacgrkrr 733
```

RESULT   3

Y01301

Y01301 standard; Protein; 733 AA.

XX

AC   Y01301;

XX

DT   07-JUN-1999   (first entry)

XX

DE   Amino acid sequence of synthetic human tropoelastin SHEL.

XX

KW   Tropoelastin; hTE; elastin; glycosaminoglycan; GAG-binding; medical;
KW   pharmaceutical; veterinary; cosmetic application; anti-wrinkle; food;
KW   hand lotion; surgical implant; industrial product; human; SHEL.

XX

OS   Synthetic.
OS   Homo sapiens.

XX

PN   WO9903886-A1.

XX

PD   28-JAN-1999.

XX

PF   17-JUL-1998;   98WO-AU00564.

PR   18-JUL-1997;   97AU-0008117.

XX

PA   (UNSI) UNIV SYDNEY.

XX

PI   Weiss AS;

XX

DR   WPI; 1999-132162/11.

DR   N-PSDB; X27704.

XX

PT   New derivatives of human tropoelastin - with elastin-like or
PT   macromolecular binding properties, useful e.g. as surgical implants

XX

PS   Disclosure; Fig 1; 82pp; English.

XX

CC   The invention relates to a derivative or variant of human tropoelastin
CC   (hTE) having elastin-like and/or macromolecule (specifically
CC   glycosaminoglycan (GAG))-binding properties. Cells containing vectors
CC   comprising the nucleic acids encoding the variants or derivatives are
CC   used to produce the proteins recombinantly. The tropoelastin derivatives
CC   or hybrid proteins containing the derivatives are useful in medical,
CC   pharmaceutical, veterinary and cosmetic applications, e.g. as anti-
CC   wrinkle or hand lotions, also as surgical implants, foods and industrial
CC   products. The hybrid protein have controllable GAG-binding properties,
CC   depending on presence or absence of a specific fragment, designated
CC   peptide 26A, from hTE. The present sequence represents the amino acid
CC   sequence of the synthetic human tropoelastin SHEL.

XX

SQ   Sequence   733 AA;

```
Query Match             100.0%;  Score 3785;  DB 20;  Length 733;
Best Local Similarity   100.0%;  Pred. No. 3e-213;
Matches  731;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1  GSYPGAIPGGYPGGVYPYPGSAGLGAGLGGGALGPGGKPLKPYPGGLAGLGAGLGAGLGAFPAVT  60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3  ggvpgaipggYpggvYfypgaglgaglggalgpggkpykplpyvpgglagagigag-gaifpavt  62

Qy   61  FPGALVPGGVADAAAAYKAAKAAGAGLGGVPGVGGLCVSAGAVVPQPGAGVKPGKVPGVGL       120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63  fpgalvpggvadaaaaykaakaagaglggvpgvggigvvsagavvpqpgqpgygagvpgkvpgrgi   122

Qy  121  PGVGPGVLPGARFPGGVLPGVPTGKLPYGYGPGPGVAGAAGKAGYPGTGVGPQAAAAAARARAKF  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123  pgvgpgvlpgarifpgvvglpgvpgtgagykpkapgvggafagipgvgpfggpgpgvplgy       182

Qy  181  PIKAPKLPGGYGLPYTTGKLPYGYGPGGVYAGRAGKAGYPGTGVGPQAAAAAAAAKAAAKF       240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183  pikapklpggygylpyttgkipyvyvgypgygyvpgaaaagkaaqypgtgvgpqaaaaaaaaakf   242

Qy  241  GAGRAGYLPGVGGAGVPGPGAGADEGVRRSLSPELRSGDPSSSQHLPSTPSSPRVPGALAAAKAA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  243  gagaagvlpgvygagvpgvgvpgagipgiggiagvgtpaaaaaaaaaaakaakygaaaglvpgg    302

Qy  301  PGFGPGVGVVPGAGVPGVGVGPGAGIPVVPGAGIPGAAVPGVVSPEARAAKAAAAKYGAR        360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  303  pgfgpgvvvvpgagvpgvgvgpgagipvvpgagvpgaavpgvvspeaaakaaakaaKygar       362

Qy  361  PGVGVGGIPYTVGVGAGGFPGFGVGVGGIPVAGVPSVGGVPGVGVGVGVGVGISPERAQAAAA     420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  363  pgvgvggipYtvgvgaggfpgfgvgvggipvagvpsvggvpgvgvgvgvgvgvgispeaqaaaa    422

Qy  421  AKAAKYGVSTPAAAAKAAAKAAQFGLVPGVGVGVAPGVGLAPGVGVAPG                   480
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db  423  akaakygvstpaaaaakaaakaaqfgivpgvgvgvgvapgvgvapgvglapgvgvapg           482

Qy  481  VGVAPGVGVAPGIGPGGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGPGLGVGA         540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  483  vgvapgvgvapgigpggvaaaaaksaakvaakaglraaaglgagipglgvgvgvpglgvga        542

Qy  541  GVPGLGVGAGVPGFGAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAA        600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  543  gvpglgvgagvpgfgagadegvrrslspelregdpssqhlpstpsSprvpgalaaakaa         602

Qy  601  KYGAAVPGVLGGLGALGGVGIPGVGVGGIPGGVGGIPGGVPAAAAKAAKAAAAKAAKYGAAGLGGLGVG  660
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  603  kygaavpgvlggIgalggvgipgvgvggipggvvggipggga                          662

Qy  661  GLGVPGVGGLGIPPAAAAKAAKYGAAGLGGVLGAGQFPLGGVAARPFGLSPIFPGGA           720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  663  glgvpgvggIgippaaaaKaaKygaagfgivgalggvggiggvg                        662
```

```
Db   663 glgvpgvrvgglgqippaaaakaakygaaglggvlggagqfplggvaarpgfglspifpgga 722

Qy   721 CLGKACGRKRK 731
         |||||||||
Db   723 clgkacgrkrk 733

RESULT 4
W46315
ID   W46315 standard; protein; 730 AA.
XX
AC   W46315;
XX
DT   23-JUL-1998 (first entry)
XX
DE   Human elastin containing non-natural polypeptide MFU-1 sequence.
XX
KW   MFU-1; minimal functional unit; elastin; human; fibrous protein;
KW   beta-sheet; coating; wound dressing.
XX
OS   Homo sapiens.
XX
     Key              Location/Qualifiers
     Protein          374..499
                      /note= "MFU-1 polypeptide"
XX
PN   WO9805585-A2.
XX
PD   12-FEB-1998.
XX
PF   07-AUG-1997; 97WO-CA00560.
XX
PR   07-AUG-1997; 97US-0911364.
PR   07-AUG-1996; 96US-0023552.
XX
PA   (HOSP-) HOSPITAL FOR SICK CHILDREN.
PA   (PROT-) PROTEIN SPECIALTIES LTD.
XX
PI   Keeley FW, Rothstein A, Rothstein SJ;
XX
DR   WPI; 1998-145551/13.
XX
PT   New non-natural polypeptide with multiple beta-sheet, beta-turn
PT   structures - particularly based on human elastin, useful for coating
PT   prostheses, as wound dressings, etc., allows ingrowth of cells
XX
PS   Claim 5; Fig 1B; 39pp; English.
XX
CC   This represents the human elastin sequence containing the minimal
CC   functional unit (MFU)-1 polypeptide of the invention. This MFU-1 is a
CC   polypeptide that has at least 3 beta-sheet/ beta-turn structures, but
CC   is not a naturally occurring fibrous protein. Each beta-sheet structure
CC   has 3-7 (preferably 5-7) amino acids and the MFU polypeptide may include
CC   at least 1 amino acid that can take part in crosslinking. The polypeptide
CC   can also be derived from the sequences of animal elastin, lamprin and
CC   spider silk protein. The MFU polypeptides are self-aligning peptides
CC   having the same primary structure as part of a natural fibrous protein.
CC   They are used to coat prostheses made of animal or synthetic material or
CC   metal, particularly for use as blood vessel or heart valve replacements,
CC   wound or burn dressings, or stents. They can be used in cosmetic,
CC   elastic or high-tensile strength materials, e.g. ropes or parachute cord.
CC   Prostheses based on the MFU allow penetration of endothelial cells, so
CC   become permanent, living, tissue replacements. The MFU polypeptides have
CC   better biocompatibility than known elastin-based materials. They are well
CC   defined, homogeneous material and are easier to manipulate and produce
CC   than full-length elastins. They are non-thrombogenic and non-immunogenic.
CC   Materials can be made from 2 or more different MFU polypeptides to allow
CC   properties to be tailored for particular applications, e.g. combining the
CC   high extensibility of elastin and the high tensile strength of spider
CC   silk protein.
XX
SQ   Sequence 730 AA.
```

```
Query Match             99.6%;  Score 3768.5;  DB 19;  Length 730;
Best Local Similarity   99.9%;  Pred. No. 2/7e-212;
Matches 730;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   1;

Qy     1 GGVPGAIPGGVPGGVFTPGAGLGAGLGGGALGPGGKPLKPVPGGLAGAGLGAGLGAPPAvt   60
         |||||||||||  ||||||  |||||| |||  |||||||||||||||| |||| |||||
Db     1 ggvpgaipggvrpggvffpgaglgalggagliggskplkpvpggllagag-lgagigaifpavt   60

Qy    61 FPGALVPGGVADAAAAYKAKAGAGLGCVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL   120
         ||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||| ||
Db    61 fpgalvpggvadaaaaykaakagaglggvpgvgglgvsagavvpqpgagvkpgkvpgygl   120

Qy   121 PGVYPGGVLPGARPGVGVLPGVTPYGAGVKPKAPGVGGAFAGIPGVGPGGPGPGVPGLGY   180
         |||||||||||| |||||||||||||||||||||||||||||||||| |||| |||||||
Db   121 pgvypggvlpgarfpgvgvlpgvtpygagvkpkapgvggafagipgvgpgpgvgpvplgy   180

Qy   181 PIKAPKLPGGYGLPVTTGKLPYGVPDGGVGAGAGKAGYPTGTGVGPQAAAAAAAKAAAKF   240
         ||||||||||||||| ||||||| | |||| |||||||||||||||||||||||||||||
Db   181 pikapklpggyglpyttgklpyygvppggvagaagkagyptgtgvgpqaaaaaaakaaakf   240

Qy   241 GAGAAGVLPGVGGAGVPGGVPGAIPGIGGIAGVGTPAAAAAAAAAKAAKYGAAAGLVPGG   300
         ||||||||||||  |||||| ||| ||||| |||| ||||||||||| |||||||||||||
Db   241 gagaagvlpgvggvpggvfpgigglagvygtpaaaaaaaaaakaaakygaaaglvpgg   300

Qy   301 PGFGPGVGVPGAGVPGAGTPGVGVGPGAGIPVVPGAGIPVVPGAAVPGVVSPEAAAKAAAKYGAR   360
         |  ||||||||||||| |||| |||| ||||||||| |||||||| ||||||||||||  |||||
Db   301 psfgpgvvvpgagvpgvgvpggagipvvpgagipgaavpgvvspeaaakaaaakygar   360

Qy   361 PGVGVGGIPTIVGVGAGGFPGFGVGNVGIPGVAGVPSVGGVGGVPGCVGISPEAQAAAA   420
         |||| |||||  ||| ||||||||||  || ||||||||||| | ||| |||||||||||
Db   361 pgvvggipt-vgvvgagfpgfgvgvvggipgvagvpsvggvgvvpvvvispeqaaaa   420

Qy   421 AKAAKYGVGTPAAAAAKAAAKAAQFGLVPGVGVAPGVGVAPGVGLAPGVGVAPG   480
         |||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   421 akaakygvgtpaaaaakaaakaaqfglvpgvgvapgvgvapgvgvlapgvgvapg   480

Qy   481 VGVAPGVGVAPGIGPGGVAAAAKSAAKYAAKAQLRAAAGLGAGIPGLGVGVGVPGLGVGA   540
         ||||||||||||||||||||||||| || ||||||||||||||||||||||||||||||||
Db   481 vgvapgvgvapgigpggvaaaaksaaakvaakaqlraaaglgagipglgvgvgvpglgvga   540

Qy   541 GVPGLGVGAGVPGFGAGADEGVRRSLSPELRESGDPSSSOHLPSTFSSPRVPGALAAKAA   600
         ||||||||||||||||||||||||||||||||| ||| ||||||| ||||||||||| ||
Db   541 svpglgvgagvpgfgagadegvrrslspelregdpssqhlpstpssprvpgalaaakaa   600

Qy   601 KYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAKAAAAKAAQGLVGAAGLGGLGVG   660
         |||||||||||||||||||| |||||||  |||||||||| ||||| |||||||| ||||
Db   601 kygaavpgv-lgglgalggvglipggvvgvgagpaaaaaakaaakaaqfglvgaaglggigvvg   660

Qy   661 GLGVPGVGGLGGIPPAAAAKAAKYGAAGLGGVLGGAGGPGLGVAARPGFGLSPIFPGGA   720
         |||||||||| |||| ||||||||||||||||||| |   ||||||||||||||||||||
Db   661 glgvpgvggi-ggipgaaaakaakygaaglggvligga.qfpigvvaarpgfglspifpgga   719

Qy   721 CLGKACGRKRK 731
         |||||||||||
Db   720 clgkacgrkrk 730

RESULT 5
Y01302
ID   Y01302 standard; protein; 698 AA.
XX
AC   Y01302;
XX
DT   07-JUN-1999 (first entry)
XX
DE   Human tropoelastin variant SHELdelta26A.
XX
KW   Tropoelastin; hTE; elastin; glycosaminoglycan; GAG-binding; medical;
KW   pharmaceutical; veterinary; cosmetic application; anti-wrinkle; food;
KW   hand lotion; surgical implant; industrial product; human; SHEJ, variant.
XX
OS   Homo sapiens.
```

```
OS   Synthetic.
XX
PN   WO9903886-A1.
XX
PD   28-JAN-1999.
XX
PF   17-JUL-1998;  98WO-AU00564.
XX
PR   18-JUL-1997;  97AU-0008117.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Weiss AS;
XX
DR   WPI; 1999-132162/11.
XX
PT   New derivatives of human tropoelastin - with elastin-like or
     macromolecular binding properties useful e.g. as surgical implants
XX
PS   Claim 13; Fig 2; 82pp; English.
XX
CC   The invention relates to a derivative or variant of human tropoelastin
CC   (hTE) having elastin-like and/or macromolecule (specifically
CC   glycosaminoglycan (GAG))-binding properties. Cells containing vectors
CC   comprising the nucleic acids encoding the variants or derivatives are
CC   used to produce the proteins recombinantly. The tropoelastin derivatives
CC   or hybrid proteins containing the derivatives are useful in medical,
CC   pharmaceutical, veterinary and cosmetic applications e.g. as anti-
CC   wrinkle or hand lotions, also as surgical implants, foods and industrial
CC   products. The hybrid protein have controllable GAG-binding properties,
CC   depending on presence or absence of a specific fragment, designated
CC   peptide 26A, from hTE. The present sequence represents the synthetic
CC   human tropoelastin variant SHEdelta26A.
XX
SQ   Sequence 698 AA;

Query Match             94.8%;  Score 3588.5;  DB 20;  Length 698;
Best Local Similarity   95.5%;
Matches  698;  Conservative    0;  Mismatches    0;  Indels   33;  Gaps    1;

Qy        1  GGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPGGLAGNGLGAGLGAFPAVT  60
             ||||||||||||||| ||||||||||| ||||||||||||||||| | ||||||||||||
Db        1  ggvpgaipggvpggvlypgaglgalgggalgpggkplkpvpggla?agl?aglgafpavt  60

Qy       61  FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL  120
             |||||||||||||||||||||||||| | | |||||||||||||||||||||||||||||
Db       61  fpgalvpggvadaaaaykaakagagl?gvp?vgglgvsagavvpqpgagvkpgkvpgvgl  120

Qy      121  PGVYPGGVLPGARFPGVGVLPGVPTGAGVKPKAPGVGGAFAGIPGVGPFPGGPQPGVPLGY  180
             ||||||||||||||||||||||  ||||||||||||||||||||  ||||||||||||||
Db      121  pgvypggvlpgarfpgvgvlpgvgvgvkpkapgvggafagipgvgvgpfpggpqpgvplgy  180

Qy      181  PIKAPKLPGGYGLPFTTGKLPYGYGPGGVAGGAAGKAGYPTGTGVGPQPAAAAKAARAKF  240
             ||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db      181  plkapklpggyglpyttgklpyygyppgyqgqaagkagyptgtgvqpqaaaaakaaakf  240

Qy      241  GAGAAGVLPGVGGAGVPGVPGGAIPGIGGGIPGVGTPAAAAAAKAAKYGAAAGLVPGG  300
             ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      241  gagaagvlpgvggagvpgvpggaipgigggipgvgtpaaaaaaakaaakygaaglvpgg  300

Qy      301  PGFPGPGTVGGVPGCAGVPGVCVPGAGIPVVPGAGIPGAAVPGGVVSPEAAAKAAKRAKYGAR  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301  p?g?p?gyvgvpgagvpgvpgvpgagipvvpggagipgaavpgvvspeeaakaaaaakaazygar  360

Qy      361  PGVGVGIPTYGVGAGFPGPGVGVGGIPGAYGVPGSVGGVPGVGGVPGVGISPEAAQAAAA  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361  pgvgvggiptygvgaggfpgpgvgvggipgayvgvpgvgvvgvpgvgvgvpgvgispeeqaaaa  420

Qy      421  AKAAKYGVGTPAAAAAAKAAAKAACFGLVPGVGVAPGGVGVAPGGVGVAPGVGVAPG  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421  akaakyqvgtpaaaaaakaakaaqfglvpgvgvapgvgvapsygvvrapgvglapgygvapg  480

Qy      481  VGVAPGVGVAPGIGPGGVAPAAAKSAAKVAAKAQLRAPAGLGAGIPGLGVGVGVPGLGVGA  540
             ||||||||||  ||||||||||||||||||  ||||||  |||||||||||||||||||||
Db      481  vgvapgvgvapsipggvaaaksaakvaaxaqlraaaglgaglpglgvgvgvpggigvgvga  540

Qy      541  GVPGLGVGAGVPGPGACADEGVRRSLSSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAA  600
             |||||||||||||||||                      |||||||||||
Db      541  gvpglgvggavpgiga-----------------------vpgalaaakaa  567

Qy      601  KYGAAVPGVLGGLGALGGVGIPGGVVGGAGPAAAAAAKAAAKAAQFGLVGAAGLGGLGVG  660
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      568  kygaavpgvlgglgalggvgipggvvggagpaaaaaakaaakaaqfglvgaaglgglgvg  627

Qy      661  GLGVGVGVGGLGGIPGAAAAKAAKYGAAGLGGVLGGAGQFPLGGVAARPGVGLSPIFGGA  720
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      628  glgvgvgvgglgggipgaaaakaakygaaglggvlggagqfplggvaarpgvglspifgga  687

Qy      721  CLGKACGRKRK  731
             |||||||||||
Db      688  clgkacgrkrk  698


RESULT  6
Y69069
ID   Y69069 standard; Protein; 698 AA.
XX
AC   Y69069;
XX
DT   30-MAY-2000  (first entry)
XX
DE   Amino acid sequence of a human reduced tropoelastin derivative.
XX
KW   Tropoelastin; derivative; SHEL-delta-26a; SHEL; proteolysis; protease;
KW   antiwrinkle; hand lotion; bulking agent; chemotaxis; proliferation;
KW   growth inhibition; peptidomimetic; lung damage; elastin; cancer;
KW   metastasis; blood clotting.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO20004043-A1.
XX
PD   27-JAN-2000.
XX
PF   19-JUL-1999;  99WO-AU00580.
XX
PR   17-JUL-1998;  98AU-0004723.
XX
PA   (UNSY ) UNIV SYDNEY
XX
PI   Weiss AS;
XX
DR   WPI; 2000-182399/16.
DR   N-PSDB; 261144.
XX
PT   New derivatives of tropoelastin useful, e.g. for inducing chemotaxis
PT   and cell growth -
XX
PS   Disclosure; Page 110-112; 136pp; English.
XX
CC   The present sequence represents a human reduced tropoelastin derivative,
CC   designated SHEL-delta-26a. The sequence is produced by removing
CC   exon 26a of SHEL (SHEL not defined). The protein is representative of
CC   tropoelastin derivatives of the invention, in which a subsequence has
CC   been mutated so that susceptibility to proteolysis is reduced or
CC   eliminated, or a subsequence has been inserted so that susceptibility
CC   to proteolysis is increased. The derivatives have with reduced
CC   susceptibility, and can be used where the wild-type protein would be
CC   degraded too easily, e.g. in contact with serum or wound exudate. The
CC   tropoelastin derivatives provide competitive inhibition of protease
CC   activity. The tropoelastin derivatives, and other polypeptides containing
CC   tropoelastin derivative-derived protease-susceptibility sites, are useful
CC   in human or veterinary medicine, cosmetics (e.g. antiwrinkle of hand
CC   lotions), as bulking agents and for inducing chemotaxis. They are also
```

CC   useful for proliferation or growth inhibition, particularly of smooth
CC   muscle cells, epithelial or endothelial cells, fibroblasts, osteocytes,
CC   chondrocytes and platelets. Peptidomimetics that mimic the protease
CC   cleavage site in tropoelastin derivatives are competitive inhibitors of
CC   the protease, and are used for protecting against lung damage caused by
CC   elastin, for inhibiting or controlling localized growth of cancers or
CC   metastases, or to limit protease activity that causes blood clotting.
XX
SQ   Sequence   698 AA;

Query Match   94.8%; Score 3588.5; DB 21; Length 698;
Best Local Similarity 95.5%; Pred. No. 7.9e-202;
Matches 698; Conservative   0; Mismatches 0; Indels 33; Gaps   1;

Qy   1 GGVPGAIPGGVPGGVFYPGAGLGGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ggvpgaipggvpggvfypgaglgalgggalgpggklkpvpggglagaglgaglgafpavt 60

Qy   61 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 fpgalvpggvadaaaaykaakagaglggvpgvggigvvpgvsagavvpqpgagvkpgkvpgvgl 120

Qy   121 PGVPGGVLPGARFPGVGVLPGVPTGAGVKPKAPGVGGAFAGIPGVGPGGCPGPGGVPLGY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 pgvpggvlpgarfpgvgvlpgvptgagvkpkapgvggafagipgvgpgqpgpgvplgy 180

Qy   181 PIKAPKLPGGYGGLPTTGKLPYGYGPGGVAGAGKGAGYPGTGVGPQAAAAAKAAAKF 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 pikapklpggyggglpttgklpygygpggvagagkgagypgtgvgpqaaaaaakaaakf 240

Qy   241 GAGAAGVLPGVGGAGVPGVPGAIPGIGGLRGVGTPAAAAAAAAKAAKYGAAAGLVPGG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 gagaagvlpgvggagvpgvpgaipgiggligygtpaaaaaaaaaaakaaagykgaaaglvpgg 300

Qy   301 PGFGPGVVGVPGAGVPGVGVPGVGVPVVPGAGIPGAAVPGVVSPEAAAKAAKAAKYGAR 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 pgfgpgvvgvpgagvpgvgvpgvgvpvvpgagipgaavpgvvspeaaakaaakaakygar 360

Qy   361 PGVGVGGIPGVGVGAGGPFGPGVGVGGIPGVAGVPGVSVGVGVGGIPGVGVPGVVPGVGGGIPGVGTVPGG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 pgvgvggipgvgvgagggipgfgvgvvgggfpgfgvvgvgvpgvsvgvpgvsgvpgvspeagaaa 420

Qy   421 AKAAKYGVGPAAAAKAAKAAQFGLVPGVGVAPGVGLAPGVGVAPGG 480
           ||||||||||||||||||||||||||||||||||||||||||||
Db   421 akaakygvgpaaaaakaaakaaqfglvpgvgvapgvgvapgvsvgvpgvsvgvapg 480

Qy   481 VGVAPGVVAPGIGPGGVAAAAKSAKVAAAAKsaakvaakaqiraaaglpgigyvgvpglgvga 540
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   481 vgvapgvvapgigpggvaaaaksakvaaaaksaakvaakaqiraaaglpgigyvgvpglgvga 540

Qy   541 GYPGLGVGAGVPGPGEAGADDGVRRSLSPEIRRGDPSSSQHLPSTPPSSPRVPGALAAAAKAA 600
           ||||||||| ||||||||||
Db   541 gypglgvgagvpgfga------VPgalaaakaa 567

Qy   601 KYGAAVPGVLGLGALGGVGVGIPGGVTVGAG 660
           |||||||||||||||||||
Db   568 kygaavpgvigglggvgvfipggvifpggvigaglggalgggigygaagigglggivg 627

Qy   661 GLGVPGVGGLGGIPPAAAKAAKKYGAAGLSGVLGGAGCQFPLGGVAARPGFGLSPIFPGGA 720
           |||||||||||||||||||||||||||
Db   628 glgvpgvggigglgggippaaaaakaaakygaaagigsvlggagigslpggacifsplifpgga 687

Qy   721 CIGKACGKKRK 731
           |||||||||||
Db   688 cigkacgrkrk 698

RESULT 7
B08630
ID   B08630 standard; peptide; 712 AA.
XX
AC   B08630;

XX
DT   20-DEC-2000 (first entry)
XX
DE   Amino acid sequence of a human elastin polypeptide.
XX
KW   Tropoelastin; elastin; elastic fibre; smooth muscle cell proliferation;
KW   smooth muscle cell differentiation; smooth muscle cell migration;
KW   smooth muscle cell function; atherosclerosis; restenosis; aneurysm;
KW   vascular bypass graft stenosis; transplant arteriopathy; dissection;
KW   SVAS; hypertension; transplant arteriopathy.
XX
OS   Homo sapiens.
XX
XX   WO200050068-A2.
PN
XX
PD   31-AUG-2000.
XX
PF   28-FEB-2000; 2000WO-US02526.
XX
PR   26-FEB-1999; 99US-0258217.
XX
PA   (UTAH) UNIV UTAH RES FOUND.
XX
PI   Keating MT, Li DY;
XX
DR   WPI; 2000-533134/48.
XX
PT   Elastin based compositions useful for treating atherosclerosis;
PT   restenosis, vascular bypass graft stenosis, transplant arteriopathy,
PT   aneurysm, dissection SVAS and/or hypertension -
XX
PS   Example 3; Page 46; 79pp; English.
XX
CC   The present sequence represents a human elastin. Peptides derived
CC   from elastin are used in compositions of the invention. The
CC   specification describes elastin based compositions that are potent
CC   regulators of smooth muscle cell proliferation, differentiation and
CC   migration in vivo. The elastin-based compositions comprise at least one
CC   elastic fibre, elastins, tropoelastins (or fragments of them) which have
CC   biological activities comprising: inhibiting the proliferation of smooth
CC   muscle cells in vivo; stimulating the differentiation of smooth muscle
CC   cell in vivo; and regulating the migration of smooth muscle cells in
CC   vivo. The compositions may be used for the prophylaxis or treatment
CC   of a disorder characterized by diminished capacity to regulate smooth
CC   muscle cell function such as atherosclerosis, restenosis, vascular
CC   bypass graft stenosis, transplant arteriopathy, aneurysm and/or
CC   dissection. Disorders which may be treated also include SVAS (undefined),
CC   hypertension, and transplant arteriopathy.
XX
SQ   Sequence   712 AA;

Query Match   91.4%; Score 3458.5; DB 21; Length 712;
Best Local Similarity 92.1%; Pred. No. 3e-194;
Matches 679; Conservative   0; Mismatches 1; Indels 57; Gaps   3;

Qy   1 GGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27 ggvpgaipggvpggvfypgaglgalggalgggalgpggkplkpvpgggagagigagigaafpavt 86

Qy   61 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87 fpgalvpggvadaaaaykaakagaglgggigvvgyvsagavvpqpgagvkpgkvpgvgl 146

Qy   121 PGVPGGVLPGARFPGVGVLPGVPTGAGVKPKAPGVGGAFAGIPGVGPFGGPQPGVPLGY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   147 pgvpggvlpgarfpgvgvlpgvptgagvkpkapgvggafagipgygvpfggpgsvplgy 206

Qy   181 PIKAPKLPGGYGGLPYTTGKLPYGYGPGGVAGAGKGAGYPGTGVGPQAAAAAKAAAKF 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   207 pikapklpggyggglpyttgklpygygpggvagagaagkasypgtgvvypgaaaaakaaakf 266

Qy   241 GAGAAGVLPGVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAKAAKYGAAAGLVPGG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 gagaagvlpgvggagvpgvpgaipgiggiagvgtpaaaaaaaaaakaaagykgaaaglvpgg 300

XX
PS   Example 3; Page 48; 79pp; English.
XX
CC   The present sequence represents a fusion protein, comprising human
CC   elastin and c-myc, preceded by a His tag. The protein is used in
CC   compositions of the invention. The specification describes elastin
CC   based compositions that are potent regulators of smooth muscle cell
CC   proliferation, differentiation and migration in vivo. The elastin-based
CC   compositions comprise at least one elastic fibre, elastins, tropoelastins
CC   (or fragments of them) which have biological activities comprising:
CC   inhibiting the proliferation of smooth muscle cells in vivo; stimulating
CC   the differentiation of smooth muscle cell in vivo; and regulating the
CC   migration of smooth muscle cells in vivo. The compositions may be used
CC   for the prophylaxis or treatment of a disorder characterized by
CC   diminished capacity to regulate smooth muscle cell function such as
CC   atherosclerosis, restenosis, vascular bypass graft stenosis, transplant
CC   arteriopathy, aneurysm and/or dissection. Disorders which may be treated
CC   also include SVAS (undefined), hypertension, and transplant arteriopathy.
XX
SQ   Sequence   730 AA;

Query Match   91.4%; Score 3458.5; DB 21; Length 730;
Best Local Similarity   92.1%; Pred. No. 3e-194;
Matches 679; Conservative   0; Mismatches   1; Indels   57; Gaps   3;

Qy   1 GGVPGAIPGGVPGGVFYPGAGLGALGALGPGGKPLKPVPGGLAGAGLGAGLGARPAVT   60
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   36 ggvpgaipggvpggvfypgaglgalgggalgpggkpllkpvpgglagaglgagipgavt   95

Qy   61 FPGALVPGGVADAAAAKKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL 120
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   96 fpgalvpggvadaaaaykaakagagiggvpgvggigvsagavvpqpgagvkpgkvpgvgl 155

Qy   121 PGVYPGGVLPGARPPGCVVLPGVPGGAGVKPAAPCVGGAFAGIPGCVGPPGGPQPGCVPLGY 180
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   156 pgvypggvlpgartpgvvgvlpgvtpgsagvxpkapgvvggafacipgvvgpfigpapgvypigv 215

Qy   181 PIKAPKLPGGYGLPYTGKLPYGVPGPGGVAAGAGKAGYPPGTGVGPQAAAAAKAAAKF   240
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   216 pikapxlpggyglpytgklpygvpgggvppvgysagvkagypptgtgvgpqaaaaakaaakf   275

Qy   241 GAGAAGYLPGVGGAGVPGVPGAIPGIGCIAGVGTPAAAAAAAAAAAKYGAARGLVPGG   300
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   276 gagaagvlpgvgggagvpgvpgaipgiggiagvgtpaaaaaaaaaaakyagaaagivpgg   335

Qy   301 PGFGPGVVGVPGGAGVPGVGVGPGAGIPVVPGAGIPVVPGGAGYVVSPEAAAKAAAAYGAR   360
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   336 pgfgpgvvgvpggagvpgvgvgpgvgvpggagipvvpggagivvpeeakkaaakaaakygar   395

Qy   361 PGVGVGGIPYGVGAGGFPGFGGVGGIPGVAGVYPGVGGIPGVYPGVPGYPGVGGVY   420
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   396 pgvgvggipyvggaggfpgfgvgggipfgvgvggipgvagvypgvpgyvgvggispeagdeaa   455

Qy   421 AKAAKYGVGTPAAAAAAKAAAKAAQF------GLVPGVGVAPGVGVAPGVGVGLAPG   474
           ||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   456 akaakygvgtpaaaaakaaakaaqfallnlaglvpgvgvapgvgvapgvgvapgylapg   515

Qy   475 VGVAPGVGVAPGVGVAPGIGPGGVAAAKSSAKVAAKAQLRAAGLGAGIPGLGVGVGVP   534
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   516 vgvapgvgvapgvgapgigpggvgvaaaaksaakvaaakglraaagigagipglgvgvgvp   575

Qy   535 GLGVGAGVPGLGVGAGVPGPGAGADEVRRSLSPELREGDPSSQHLPSTPSSPRVPGAL   594
           |||||||||||||||||||||||||||||||||||||||||
Db   576 glgvgagvpglgvgagvpgfga-------------------------vpgal   602

Qy   595 AAKAAAKYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAKAAAKAAQFGLVGAAGL   654
           |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   603 aaakaaakygaavpgvlggigalggvgipggvvgagpaaaaaaakaaakaaqfglvgaagl   662

Qy   655 GGLGVGGLGVPGGVGGLGGLGGiPPAAAAAKAAKYGAAGLGGVLGGAGQFPLGGVAARPGFGLSP   714
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   663 gglgvgglgvpggvggiggippaaaaakaaky--------------gvaarpgfglsp   704

RESULT   8
B08631
ID   B08631 standard; peptide; 730 AA.
XX
AC   B08631;
XX
DT   20-DEC-2000   (first entry)
XX
DE   Fusion protein comprising human elastin and c-myc.
XX
KW   Tropoelastin; elastin; elastic fibre; smooth muscle cell proliferation;
KW   smooth muscle cell differentiation; smooth muscle cell migration;
KW   smooth muscle cell function; atherosclerosis; restenosis; aneurysm;
KW   vascular bypass graft stenosis; transplant arteriopathy; dissection;
KW   SVAS; hypertension; transplant arteriopathy.
XX
OS   Synthetic.
OS   Homo sapiens.
OS   Unidentified.
XX
PN   WO2005068-A2.
XX
PD   31-AUG-2000.
XX
PF   28-FEB-2000; 2000WO-US02526.
XX
PR   26-FEB-1999;   99US-0258217.
XX
PA   (UTAH) UNIV UTAH RES FOUND.
XX
PI   Keating MT, Li DY;
XX
DR   WPI; 2000-533134/48.
XX
PT   Elastin based compositions useful for treating atherosclerosis,
PT   restenosis, vascular bypass graft stenosis, transplant arteriopathy,
PT   aneurysm, dissection SVAS and/or hypertension -

Db   267 gagaagvlpggagvpggvpgaipgiggiagvgtpaaaaaaaaaaakygaaaglvpgg   326
           ||||||||||||||||||||||||||||||||||||||||||||||
Qy   301 PGFGPGVVGVPGGAGVPGVGVGPGAGIPVVPGAGIPVVPGG   360
           |||||||||||||||||||||||||||||||||||||||||
Db   327 pgfgpgvvgvpggagvpgvgvgpgvgvpggagipvvpggagipvvpggar   386

Qy   361 PGVGVGGIPYGVGAGGFPGFGGVGGIPGVAGVYPGVGGIPGVYPGVPGYPGVGGVY   420
           |||||||||||||||||||||||||||||||||||||||||
Db   387 pgvgvggipyvggaggfpgfgvgggipfgvgvggipgvagvypgvpgyvgvggispeagdeaaa   446

Qy   421 AKAAKYGVGTPAAAAAAKAAAKAAQF------GLVPGVGVAPGVGVAPGVGVGLAPG   474
           |||||||||||||||||||||||||| ||||||||||
Db   447 akaakygvgtpaaaaakaaakaaqfallnlaglvpgvgvapgvgvapgvgvapgylapg   506

Qy   475 VGVAPGVGVAPGVGVAPGIGPGGVAAAKSSAKVAAKAQLRAAGLGAGIPGLGVGVGVP   534
           |||||||||||||||||||||||||||||||||||||||||
Db   507 vgvapgvgvapgvgvapgvgapgigpggvgvaaaaksaakvaaakaqlraaagigagipglgvgvgvp   566

Qy   535 GLGVGAGVPGLGVGAGVPGPGAGADEVRRSLSPELREGDPSSQHLPSTPSSPRVPGAL   594
           |||||||||||||||||||||||||||||||||||||||| vpgal
Db   567 glgvgagvpglgvgagvpgfga-------------------------vpgal   593

Qy   595 AAKAAAKYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAKAAAKAAQFGLVGAAGL   654
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   594 aaakaaakygaavpgvlggigalggvgipggvvgagpaaaaaaakaaakaaqfglvgaagl   653

Qy   655 GGLGVGGLGVPGGVGGLGGLGGiPPAAAAAKAAKYGAAGLGGVLGGAGQFPLGGVAARPGFGLSP   714
           ||||||||||||||||||||||||||||||||||||||
Db   654 gglgvgglgvpggvggigglggippaaaaakaaky--------------gvaarpgfglsp   695

Qy   715 IFPGGACIGKACGRKKK   731
           |||||||||||||||||
Db   696 ifpggacigkacgrkrk   712

```
Qy   715 IFPGGACLGKACGRKRK 731
         |||||||||||||||||
Db   705 ifpggacigkacgrkrk 721
```

RESULT 9
YO1303
ID   YO1303 standard; Protein; 660 AA.
AC   YO1303;
XX
DT   07-JUN-1999 (first entry)
XX
DE   Human tropoelastin derivative SHELdeltamodified.
XX
KW   Tropoelastin; hTE; elastin; glycosaminoglycan; GAG-binding; medical;
KW   pharmaceutical; veterinary; cosmetic application; anti-wrinkle; food;
KW   hand lotion; surgical implant; industrial product; human; SHEL.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9903886-A1.
XX
PD   28-JAN-1999.
XX
PF   17-JUL-1998; 98WO-AU00564.
XX
PR   18-JUL-1997; 97AU-0008117.
XX
PA   (UNSY) UNIV SYDNEY.
XX
PI   Weiss AS;
XX
DR   WPI; 1999-132162/11.
DR   N-PSDB; X27705.
XX
PT   New derivatives of human tropoelastin - with elastin-like or
PT   macromolecular binding properties; useful e.g. as surgical implants
XX
PS   Claim 7; Fig 3; 82pp; English.
XX
CC   The invention relates to a derivative or variant of human tropoelastin
CC   (hTE) having elastin-like and/or macromolecule (specifically
CC   glycosaminoglycan (GAG))-binding properties. Cells containing vectors
CC   comprising the nucleic acids encoding the variants or derivatives are
CC   used to produce the proteins recombinantly. The tropoelastin derivatives
CC   or hybrid proteins containing the derivatives are useful in medical,
CC   pharmaceutical, veterinary and cosmetic applications e.g. as anti-
CC   wrinkle or hand lotions, also as surgical implants, foods and industrial
CC   products. The hybrid protein have controllable GAG-binding properties,
CC   depending on presence or absence of a specific fragment, designated
CC   peptide 26A, from hTE. The present sequence represents a human
CC   tropoelastin derivative SHELdeltamodified.
XX
SQ   Sequence 660 AA;

Query Match   85.1%; Score 3222; DB 20; Length 660;
Best Local Similarity 87.6%; Pred. No. 1.6e-180;
Matches 644; Conservative 3; Mismatches 3; Indels 80; Gaps 7;

```
Qy    1 GGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT 60
        ||||||||||||||||||||| ||
Db    2 ggvpgavpggvpggvfYpgagfga------------------------------------- 25

Qy   61 FPGALVPGGVADAAAAYKAAKGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL 120
        ||||||||||||||||||||||||||||||||||||||||||||
Db   26 ----vpggvadaaaaYkaakgagglggvpgvgglgvsagavvpgpgagvkpgkvpgvgl 80

Qy  121 PGVYPG-GVLPGGARPPGVGVLPGVPTGAGVKPKPAPGVGGAPAGIPGVGPFGGPQPGVPLG 179
        ||||||  :||||||||||||||||||||||||||||||||||||||||||||
Db   81 pgvypgfgavpggaripgvygvipgvtgagvkpgvpvytgagvkpkapgvggafagipgvygpfggpgpsvplg 140

Qy  180 YPIKAPKLPGGYGLPYTGKLPYGYGGYVAGNAGKAGYPTGTGVGPQAAAAAAAKAKAK 239
        |||||||||||||||||||| ||| ||||||||||||||||||||||
Db  141 Ypikapklpggygylpytgklpygygpgyva-aagkagvpytg-gvgpgaaaaaakakaak 199

Qy  240 FGAGRA--GVLPGVGGAGVPGVPGCAIPGIGGIAGVCTPAAAAAAAAAKAKYGAAAGLV 297
        ||||||   |  : :|||||||||||||||||||||||||||||||||||||||||||
Db  200 fgagaagfgavpgvggagvpgvpvpggaipgyiagvtgaaaaaaakakaakygaaagiv 259

Qy  298 PGGPGPGPGVVGVPGAG-VPGVGVPGAGIPVVPGAGIPGAAVPGVVSPEAAAKARAAKAK 356
        |||||||||||||||||  ||||||||||||||||| | ||||||||||||||||||||||
Db  260 pggpgfgpgvvgypgagipgavpgvvpgagipvvpgagipgaavpgvvspeaaakaaakak 319

Qy  357 YGARPGVGVGGIPTYGVGAGPGPGCVGVGGIPGVAGVPSVGGVPGVGGVPGCVGISPEAG 416
        |||||||||||||||||| |||||||||||||||||||||||| | |||||||||||||||
Db  320 Ygarpgvgvgggiptygvvgagffpgfgvgvvggipgvvggipgvagvpsvggvpgvgisngeaq 379

Qy  417 AAAAKAAKYGVGTPAAAAAKAAKAAQFGLVPGVGVVAGPGVGVAPGVGLAPGVG 476
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  380 aaaakaakygvgtpaaaaakaaakaagfgivpgvgvvapgvgvapgvglapgvg 439

Qy  477 VARGVVAPGVGVVAPGIGPGGVAAAAKSRAKVAAKAQLRAAAGLCAGIPGLGVGVGPGL 536
        |||||  ||||||||||||||||||||  |||||||||||||||||||
Db  440 vapgvvapgvgvvapgipgvggvaaaaksaakvaaaqlraaagligagipglgvgvg-gpgl 499

Qy  537 GVGAGVPGGLGVGACVPGFGAGADEGVRRSLLSPELREGDPSSSQHLPSTPSSPRVPGALAA 596
        ||||||||||                                       ||||||
Db  500 gvgagvpglgvgagvpgfga--------------------------------vpgalaa 526

Qy  597 AKAAKYGAVPGCVLGGLGALGGVGIPGGVVGAGPAAAAAAKAAKAAQFGLVGAAGLGG 656
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  527 akakyyg-avpgvlgglgalggvvglpggvvvgagpaaaaaaakaaakaaqfgivgaag-jgg 585

Qy  657 LGVGGLGVPGVGGLGGIPPAAAAAKAAKYGAAGLGGVLGGAGQFPLGGVAARPGFGLSPIF 716
        ||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db  586 lgvgglgvpgvgvglgippaaaaakaakygaagiggvlggagqfplggvaarpgfgispif 645

Qy  717 PGGACLGKACGRKRK 731
        ||||||||||
Db  646 pggacigkacgr-krk 660
```

RESULT 10
Y69071
ID   Y69071 standard; Protein; 571 AA.
AC   Y69071;
XX
DT   30-MAY-2000 (first entry)
XX
DE   Amino acid sequence of a human tropoelastin derivative.
XX
KW   Tropoelastin; derivative; proteolysis; protease; antiwrinkle;
KW   hand lotion; bulking agent; chemotaxis; proliferation;
KW   growth inhibition; peptidomimetic; lung damage; elastin; cancer;
KW   metastasis; blood clotting.
XX
OS   Homo sapiens.
XX
PN   WO20000043-A1.
XX
PD   27-JAN-2000.
XX
PF   19-JUL-1999; 99WO-AU00580.
XX
PR   17-JUL-1998; 98AU-0004723.
XX
PA   (UNSY) UNIV SYDNEY.
XX
PI   Weiss AS;
XX
DR   WPI; 2000-18239/16.

```
XX
PT   New derivatives of tropoelastin useful, e.g. for inducing chemotaxis
PT   and cell growth -
XX
PS   Disclosure; Page 115-117; 136pp; English.
XX
CC   The present sequence represents a human tropoelastin derivative, which
CC   is representative of tropoelastin derivatives of the invention. In the
CC   tropoelastin derivatives of the invention a subsequence has been
CC   mutated so that susceptibility to proteolysis is reduced or eliminated,
CC   or a subsequence has been inserted so that susceptibility to proteolysis
CC   is increased. The derivatives have with reduced susceptibility to
CC   be used where the wild-type protein would be degraded too easily,
CC   e.g. in contact with serum or wound exudate. The tropoelastin
CC   derivatives provide competitive inhibition of protease activity. The
CC   tropoelastin derivatives, and other polypeptides containing tropoelastin
CC   derivative-derived protease-susceptibility sites, are useful in human
CC   or veterinary medicine, cosmetics (e.g. antiwrinkle or hand lotions), as
CC   bulking agents and for inducing chemotaxis. They are also useful for
CC   proliferation or growth inhibition, particularly of smooth muscle cells,
CC   epithelial or endothelial cells, fibroblasts, osteocytes, chondrocytes
CC   and platelets. Peptidomimetics that mimic the protease cleavage site in
CC   tropoelastin derivatives are competitive inhibitors of the protease, and
CC   are used for protecting against lung damage caused by elastin, for
CC   inhibiting or controlling localized growth of cancers or metastases, or
CC   to limit protease activity that causes blood clotting.
XX
SQ   Sequence    571 AA;

Query Match       75.8%;   Score 2869;  DB 21;  Length 571;
Best Local Similarity   99.6%;   Pred. No. 5.1e-160;
Matches  552;   Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy       1 GGVPGAIPGGVPGGVPYPGAGLGAGLGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ggvpgaipgsgvpggvfypgsglgaglgggalgpggkplkpvpggllagaglgaglgafpavt 60

Qy      61 FPGALVPGVADAAAAYKAAKAGAGLGGVPGGVGGLGVSAGANVVPQPGAGVKPGKVPGVGL 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 fpgalvpgvadaaaaykaakagagligsvpggvgglgvsagavvpqpgagvkpgkvpgvgl 120

Qy     121 PGVVPGGVLPGARFPGVGVLPGVPTGAGVKPKPAPGVGGAFAGIPGVGPFGGPQPGVPLGY 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 pgvvpggvlpgarfpgvgvlpgvprtgagvkpkapgvggafagipgvgpfggpqpgvplgy 180

Qy     181 PIKAPKLPGGVGLPYTTGKLPYGYGPGGVAGAAGKAGYPPGCTGVGPQAAAAAAKRAAKF 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 pikapklpggvgylpyttgklpyygypgggvagaaagkagyrptgtgvgvgpqaaaaakaaakf 240

Qy     241 GAGAAGVLPGVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAKAAKYGAAGLVPGG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 gagaagvlpgvgggagvpgvpgalpgigg;agvgtpaaaaaaaaakaakygaaglvpgg 300

Qy     301 PGTPGPGVVCGVPGAGVPGVGVPGAGIPVVPGAGIPGAAYPGVVSPEAAAKAAAKAAKYGAR 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 pgfgpgvvvpgagavpgvpvpgagvpvpggipgaaypgvvspeaaakaaakaaakygar 360

Qy     361 PGVGVGGIPTYGVGAGGPFGFGVGGIPGVAGVGVGGIPGVAGVGVGGGYVGVGSPFAQAAA 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 pgvgvgggiptyvgagglpfgvgvggipgvaggipgvagvpgvsgvsvgggygvggspeaqgaaaa 420

Qy     421 AKAAKYGVGTPAAAAAKAAAKAARKAAQFGLVPGVGVGVTARGVGVCVTARPGVGLAPGVGVAPG 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 akaaxyvgtpaaaaaaakaaakaag;rglvpgvgvgvapgvgvapgvsgvgvapgvgvapg 480

Qy     481 VGVVAPGVTAPGIGPGGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGVPGLGVGA 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 vgvapgvgvapgipggsvaaaaksaaakvaaakaq;lraaagl;gacglpglgvgvpglgvga 540

Qy     541 GVPGLGVGAGVPGF 554
           |||||||||| ||
Db     541 gvpglgvgagvpgf 554
```

```
RESULT  11
Y69135
ID   Y69135 standard; Protein; 515 AA.
XX
AC   Y69135;
XX
DT   30-MAY-2000   (first entry)
XX
DE   Amino acid sequence of a human tropoelastin derivative.
XX
KW   Tropoelastin; derivative; proteolysis; protease; antiwrinkle;
KW   hand lotion; bulking agent; chemotaxis; proliferation;
KW   growth inhibition; peptidomimetic; lung damage; elastin; cancer;
KW   metastasis; blood clotting.
XX
OS   Homo sapiens.
XX
PN   WO200004043-A1.
XX
PD   27-JAN-2000.
XX
PF   19-JUL-1999;   99WO-AU00580.
XX
PR   17-JUL-1998;   98AU-0004723.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Weiss AS;
XX
DR   WPI; 2000-182399/16.
XX
PT   New derivatives of tropoelastin useful, e.g. for inducing chemotaxis
PT   and cell growth -
XX
PS   Disclosure; Page 131-133; 136pp; English.
XX
CC   The present sequence represents a human tropoelastin derivative, which
CC   is representative of tropoelastin derivatives of the invention. In the
CC   tropoelastin derivatives of the invention a subsequence has been
CC   mutated so that susceptibility to proteolysis is reduced or eliminated,
CC   or a subsequence has been inserted so that susceptibility to proteolysis
CC   is increased. The derivatives have with reduced susceptibility, and can
CC   be used where the wild-type protein would be degraded too easily,
CC   e.g. in contact with serum or wound exudate. The tropoelastin
CC   derivatives provide competitive inhibition of protease activity. The
CC   tropoelastin derivatives, and other polypeptides containing tropoelastin
CC   derivative-derived protease-susceptibility sites, are useful in human
CC   or veterinary medicine, cosmetics (e.g. antiwrinkle or hand lotions), as
CC   bulking agents and for inducing chemotaxis. They are also useful for
CC   proliferation or growth inhibition, particularly of smooth muscle cells,
CC   epithelial or endothelial cells, fibroblasts, osteocytes, chondrocytes
CC   and platelets. Peptidomimetics that mimic the protease cleavage site in
CC   tropoelastin derivatives are competitive inhibitors of the protease, and
CC   are used for protecting against lung damage caused by elastin, for
CC   inhibiting or controlling localized growth of cancers or metastases, or
CC   to limit protease activity that causes blood clotting.
XX
SQ   Sequence    515 AA;

Query Match       70.8%;   Score 2680;  DB 21;  Length 515;
Best Local Similarity   100.0%;  Pred. No. 4.7e-149;
Matches  515;   Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       1 GGVPGAIPGGVPGGVPYPGAGLGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ggvpgaipgsgvpggvfypgsglgaglgggalgpggkplkpvpggllagaglgaglgafpavt 60

Qy      61 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGGVGGLGVSAGAVVPQPGAGVKPGCKVPGVGL 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 fpgalvpggvadaaaayxaakagaglgvpggvgvgglgvsagavvpqpgagvkpgvkvpgvgl 120
```

Qy 121 PGVYPGGVLPGCARPPGCVGVLPGVPPGCGVGAFAGIPGVGPGPQPGVPLGY 180
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 pgvypgv-lpgafpgvgyvlpgtgagvlpgsagipgvgpfagggpgpgvplgy 180

Qy 181 PIKAPKLPGGIGLPYTTGKILPYGYGPGGVNAGKAGYPPGTGVGPQAAAAAKAAAAKF 240
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 181 pikapklpgsyrglpyttgkipysyygpggyvagaagkasypqtgvgpqaaaaaakaaaakf 240

Qy 241 GAGAAGVLPGVGVGGAGVPGVPGAIPGIGIAGVGTPAAAAKAAAKYGAAAGLIVPGG 300
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 gagaagvlpgvgvgvpgagipgiggipgvtpaaaaaaakaakygaaaglvvpgg 300

Qy 301 PGYGPGVGVPGGAGVPGVGVPGAGIPVVPGAGIPGAANVPGVVSPEAAAKAAAKAAKYGAR 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 pgfgpgvgvpgvgvpgaggvpgvgvpgaggipgaavpgvvspeaaakaaakygar 360

Qy 361 PGVSVGGGIPYTGYGVGAGGPPGFGVGVGGIPGVGAGVPGVGGVPGVVSPEAAQAAAA 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 pgvsvggipgtgygvgaggfpgfgvgvggipgvgvagyvpgvggvpgvvispeeqqaaat 420

Qy 421 AKAAAKYGVGPAAAAKAAAKAAAQFGLVPGVGVPGVGVAPGVGVGVGVAPGVGLAPGVGVAPS 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 akaakygvgpaaaakaaaakaaaqfglvpgvgvpgvgvapgvgvapgvgvapgvgvapgvgvaps 480

Qy 481 VGYAPGVGVAPGIGPGGVYAAAKSAAKYAAKAQLR 515
         |||||||||||||||||||||||||||||||||
Db 481 vgvapgvgvapgvgpggyyaaaksaaksaakvaaaqlr 515

RESULT 12
R80253
ID   R80253 standard; peptide; 988 AA.
XX
AC   R80253;
XX
DT   17-APR-1996 (first entry)
XX
DE   Polymer SELP7.
XX
KW   Fibroin; elastin; repeat sequence; suture; thread; pin; gel; silk;
KW   polymer; E.coli; EC3.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..76
FT   /note= "polymer repeat block sequence"
XX
PN   WO9524478-A1.
XX
PD   14-SEP-1995.
XX
PF   10-MAR-1995; 95WO-US02772.
PR   11-MAR-1994; 94US-0212237.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J;
XX
DR   WPI; 1995-328270/42.
XX
PT   Protein polymer comprising alternating blocks of fibroin and elastin
PT   units - used to form a device e.g. a suture to keep separated
PT   viable tissue together
XX
PS   Example 1; Page 27-30; 46pp; English.
XX
CC   The sequences represented by R80251-R80257 are polymers constructed of
CC   repeating blocks of fibroin-like (see R80249) and elastin-like (see
CC   R80250) units. This sequence contains 13 repeats of a block consisting
CC   of 8 elastin-like repeats followed by 6 fibroin-like repeats. The DNA

CC   sequences encoding these polymers were inserted into plasmids which were
CC   used to transform E.coli strain EC3. The polymers could then be
CC   isolated from the fermented strains by standard centrifugation
CC   techniques. The polymers are used to form a device (such as a suture,
CC   pin, thread, gel or film) to keep separated viable tissue together. By
CC   varying the the ratio of the two repetative units, and by altering the
CC   lengths of the blocks of each of them, the tensile properties of the
CC   polymer can be altered moderatly. By reducing the number of repeating units
CC   of this sequence, or by increasing the number of units of the elastin
CC   like repeat, a faster rate of resorption can be achieved. Of the
CC   polymers, SELP0 (see R80251) had the fastest resorption rate. The SELP4
CC   (see R80255) and SELP5 (see R80256) polymers showed no resorption after
CC   seven weeks. The other three polymers showed intermediate resorption.
CC   No data was given for SELP6 (see R80257).
XX
SQ   Sequence 988 AA;

Query Match   30.6%; Score 1159.5; DB 16; Length 988;
Best Local Similarity   40.5%; Pred. No. 2.8e-60;
Matches 353; Conservative 54; Mismatches 279; Indels 185; Gaps 51;

Qy   2 GVPG-AIPG-GVPG----GVYPGAGLGALGGGALGPG--------GKFLKPVPGGLAGAG 48
         |||| ||| :|||| || | | | || | | |:|||| ||| | |
Db   10 gvpgvypgvypgvypgvypgvypgvvpgvggagsgagsgagsgagsgagag 69

Qy   49 LGAGLGAFPAVFPPGALVPGGVADAAAAYKAAKAAGLG--GVPGVGLGVSAGAVVPQ- 105
         | :| | | | |: :| | ::| | | ||||| |:| || :
Db   70 sgagagsvpgvgvypgvgvypgvgvypgvgvypgvgvypgvypgvggagssgagagsg 129

Qy 106 ------------------PGAGVKPGAVPGVGLPGV-YPGGVLPGARFPGVGVLPGVPT-GAGV 149
                            ||| | | |||||| | | |:| | :|||| ||| | ||
Db 130 agagsgagsgagsgagsgagsvpgvgvypgvgvypgvypgvgvypgvgv 188

Qy 150 KPKAPGVGGAPAG------------------IPGVGPFGGPGPQPGVPL 178
         | | |||| | |||| | | | |
Db 189 ---pgvggagsgagsgagsgagsgagsvpgvgvypgvgv---pgvgyp- 240

Qy 179 GYPIKAPKLPG---GYGLPYTTGKLPYGYGPGVGAGAAGKAGYPTGTGVGPQAAAAAA 234
         | | | | :| | | ||| | | | ||| || | |:||||
Db 241 gvgvypgvgvypgvgvypgv------qvgvypgvggagssgagagsgagagsgagagsg 292

Qy 235 KAAAKFGAGAAGVLPGVG--GAGVP--GVPG-AIPGIG---GIAGVGTPAAAAAAAA 285
         | | | | | ||||| | |||| |:| | || ||||| ||||
Db 293 gagagsgag-agsvpgvgvypgvgvypgvgvypgvgvypgvypgvggagssg 351

Qy 286 KAAKYGAAAGLVPG-----GPGIGPGVGVPGAGVGVPGAGIP--VVPGAGIPGAAV 338
         | | || | || | |||||||||| ||||||| | |||||||| ||
Db 352 agagsgagsgagsgagsgagsvpgvgvypgvgvypgvgvypgvgv 411

Qy 339 PGVVSPEAAAAAKAAAKYGAARPGVGVGGIPTVGVGAGGPPGFGVGVGGIPGV---AGV 394
         ||| | | | || | |||| ||| |||| | | |||| |||| ||
Db 412 pgvypgvggagsgagsgagsgagsga---gagsgagssgagagsvpgvgvypgvgv 467

Qy 395 PSVG----GVPGVG---GVPGVGISPEAQAAAAAAKAAKYGVGTPAAAAKA-----AAK 441
         |:|| ||||| ||||| | | | | | ||| |||||| ||
Db 468 pgvgvypgvgvypgvgvypgvgv-pgvggagssgagagsgagagsgagagsgagags 526

Qy 442 AAQFGLVPGVGVAPGVGVARGVGVAPGVGLAPGVGVARGVGVAPGIGPGGVARA 501
         | | |||||||| | ||| | ||||| |||||||| | ||||||||||||
Db 527 gagagsvpgvgv-pgvgv-pgvgv-pgvgv-pgvgv-pgvggagsagsg 579

Qy 502 AKSAAKYAAKAQLRAAAGLGA-----------GIPGLGV-GVGVPGLGV-GAGVPGLGV 547
         | | : | : ||||| | | |||| :|||||| | ||||| ||
Db 580 agagsgagsgagsgagsgagsvpgvgvypgvgvypgvgvypgvgvypgvgvypgvgvypgvgv 639

Qy 548 -GAGVPGF-GAGADEGVRSLSPFLRBG-DPSSCQHLPSTPSSPRVPGALAAAKAAKYGA 604
          | |||| | | | | | | | | ||| | | | | ||
Db 640 pgvgvypvggagsgagsgagsgagsgagsgagsgagsgacsagsvpg-vgvpgvgvpgv 698

Qy 605 AVPGV-LGGLGALG----GVGIPG-GVGAGPAAAAAAKAAAKAAQFGL----------- 648
         |||| || || | :|||| :| | ||||||||||||||| |
Db 699 gvpgvgvypgvgvypgvgvypgvgvypgvgvggagsgagsgagsgagagsgagagsgaga 758

```
Qy  649  ----VGAAGLG--GLGVGGLGVPGVG--GLG----GIPPAAAAKAAKYGAAGLG------ 690
             |  ||  |.|| |.|:|||  |.|:   :|  ||  |  .| | |
Db  759  gsvpgvgvpgvgvpgvgvpgvgvpgvgvpgvgvpgvpgvvggpgagagsgagsgagags 818

Qy  691  --GVLGGAGQFPLAGVAARPFGLSPI-FPG 718
           |  ||  :|  | ||  : .|  | |:
Db  819  gagagsgagsgagsgagsvpgvgvpgvgvg 849
```

RESULT 13
R80252
ID R80252 standard; peptide; 832 AA.
XX
AC R80252;
XX
DT 17-APR-1996 (first entry)
XX
DE Polymer SELP8.
XX
KW Fibroin; elastin; repeat sequence; suture; thread; pin; gel; silk;
KW polymer; E.coli; EC3.
XX
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Peptide   1..64
FT   /note= "polymer repeat block sequence"
XX
PN WO9524478-A1.
XX
PD 14-SEP-1995.
XX
PF 10-MAR-1995; 95WO-US02772.
XX
PR 11-MAR-1994; 94US-0212237.
XX
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI Cappello J;
XX
DR WPI; 1995-328270/42.
XX
PT Protein polymer comprising alternating blocks of fibroin and elastin
PT units - used to form a device e.g. a suture to keep separated
PT viable tissue together
XX
PS Example 1; Page 25-27; 46pp; English.
XX
CC The sequences represented by R80251-R80257 are polymers constructed of
CC repeating blocks of fibroin-like and elastin-like (see R80249) and elastin-like (see
CC R80250) units. this sequence contains 13 repeats of a block consisting
CC of 8 elastin-like repeats followed by 4 fibroin-like repeats. The DNA
CC sequences encoding these polymers were inserted into plasmids which were
CC used to transform E.coli strain EC3. The polymers could then be
CC isolated from the fermented strains by standard centrifugation
CC techniques. The polymers are used to form a device (such as a suture,
CC pin, thread, gel or film) to keep separated viable tissue together. By
CC varying the the ratio of the two repetative units, and by altering the
CC lengths of the blocks of each of them, the tensile properties of the
CC polymer can altered moderatly. By reducing the number of repeating units
CC of this sequence, or by increasing the number of units of the elastin
CC like repeat, a faster rate of resorption can be achieved. Of the
CC polymers, SELP0 (see R80251) had the fastest resorption rate. The SELP4
CC (see R80255) and SELP5 (see R80256) polymers showed no resorption after
CC seven weeks. The other three polymers showed intermediate resorption.
CC No data was given for SELP6 (see R80257).
XX
SQ Sequence 832 AA;

Query Match   30.6%; Score 1159; DB 16; Length 832;
Best Local Similarity   44.4%; Pred. No. 2.6e-60;
```

Matches 350;  Conservative 49;  Mismatches 243;  Indels 146;  Gaps 55;

```
Qy  2    GVPG-AIPG-GVPG-----GVFYPGAGLGALGGGALGPGGKPIKPVPGGLAGGLGAGLGA 55
              |  |  |  |||     |  | .| :.|   | |   ||   |  |  |||:|| ||
Db  10   gvpgvgvpgvgvpgvgvpgvgvpgvgvpgvgvpgvpg----agsgagsgagagsgagags 64

Qy  56   FPAVTPGALVPGGVADAAANYKAAKAGAGLG--GVPGVGGLGVSAGAVVPQ---PGAGV 110
              | || || |||     |  | | | |  |    :| ||||| | ||  |     |:|
Db  65   vpgvgpsvgpgvgvpgvgvpgvgvpgvgvpgvpgvggagsgagsgagsgagsgsga 124

Qy  111  KPGKVPGVGLPGV-YPGGVLPGAREPGVGVLPGVPT-GAGVKPKAPGVGGAPAG------ 162
              | |||||||.||  | :|| ||   | |  || |||    |   | | |   | |
Db  125  gagsvpgvgvpgvgvpgvgvpgvgvpgvgvpgvgvpgvgvp---pgvggagsgagsgag 179

Qy  163  ---IPGVG--PFGGPQPGVP-LGYPIKAPKLPGGVGLPYTTGKLPYGVG- 205
             |||| ||  | |  ||  |  || :| |  || | |  | | |  |  |
Db  180  sgagsgagsgagsgagsvpgvgvpgvgvpgvgvpgvgvp------gvgvp- 223

Qy  206  PG-GVAGAAGKAGYPIGVGVGPLGAAAAAAAKAAAAKFGAGAAGV-LPGVG--GAGVPGVPG 261
              || |.|  | | |||| | |  |   |   ||  | | |  |  |  | |  | | |  |
Db  224  pgvgvpgvgvpgvgvpg-agsgagsgagsgagsgagsvpgvgvpgvgvpgvgvpgvgvpgv-- 280

Qy  262  AIPGIGGIAGVGTPAAAAAAAAKAAKYGAAAGLVPGGPGPG--PGVVGVPGAGVPGVG 319
             |:| | ::| | |   |  | ||  | | ||   | | | |   || | | | | |
Db  281  gvpgv-gvpgvgvpgvgvpgvggagsgagsgagsgagsgag---sgagsvpg-vgvpgvgvpgvg 335

Qy  320  VPGAGIPVVPSGAGIPGAAVPGVVGSPEAAAKAAAAKYGARPGVGVGGIPTYGVGAGGFP 379
              ||| |  | :  | |  || |  | |   |   | ||  | |||| | |  || |
Db  336  vpgvg--vpgvgvpgvgvpgvgvpgvgvpgvgvpgvggagsga----gagsgags----gagsgsvp 386

Qy  380  GFGVGVGGIPGVAGVPSVGGVPGVG----GVPGVGISPEAQAAAAAAAKAAKYGCVGSTPAAAA 435
              | | | | |  || | | | | | |    | | ||| |   | |   || | || |  | |
Db  387  gvgvpgvgvpgv-gvpgv-gvpgvgvpgvgvpgvgvgsvpgv-gvpgvgv---gagsgagsga--gagsgagsga 441

Qy  436  AKAAAKAAQFGLVPGVGVAPGVGVPSVGGVGVAPGVGVARGVGVGVGVARGIGP 495
             | | | || | | || | | |  | | | |  ||| | | | | | | ||:|:
Db  442  sgaga-----gsvpgvg-pgvgv-pgvgv-pgvgv-pgvgv-pgvgv-pgvgg 489

Qy  496  GGYAAAAASAAKYAAKAQLPAAAGLGAGIPGLGVY-GVGTPGLGV-GAGVPGLGV-GAGVP 552
              | | || |   |  | | | | | | || | | | ||||| | | | | | | | | |
Db  490  agagsgagsgagsgagsgagvpgvgvpgvgvpgvgvpgvgvpgvgvpgvgvpgvgvpgvgvp 549

Qy  553  GF-GACADEGVGRKSLSPELREGDPSSSQHLPSTPSSFRVPGALAAAKAAAKYGRAAVPGV-L 610
              | | | | | |  | | | | | |  | | | |   | | |  |   | | | | |  | |
Db  550  gvgvpgvgvpgvgvpgvgvpgvgvpgvgvpgvgvpgvgvpg---------vpg-vgvpgvgvpgvgvpgvgv 597

Qy  611  GGLGALG----GVGIPG-GVVGAGPAAAAAAKAAAAAQFGLVGRAGLGGLGVGGLGVP 665
              | | | |    | | | | | || |    |  |  || | | | | | | || | |
Db  598  pgvgvpgvgvpgvgvpgvgvpgvggagsgagsgagsgagsgagsvpgvgvpgvgvpgvgvp 657

Qy  666  GVG------GLG----GIPPAAAAKAAKYGAAGLGGVLG--GAGQFPLGGVAARPFG 711
             | |        | |    | | | | | || | | | | |    | | | || | | |
Db  658  gvgvpgvgvpgvgvpgvgvpgvggagsgagsgagsgagsgagsvpgvgvpgv---pgvg 714

Qy  712  LSPI-FPG 718
              | | |||
Db  715  vpgvgvpg 722
```

RESULT 14
R80251
ID R80251 standard; peptide; 936 AA.
XX
AC R80251;
XX
DT 17-APR-1996 (first entry)
XX
DE Polymer SELP0.
XX
KW Fibroin; elastin; repeat sequence; suture; thread; pin; gel; silk;
KW polymer; E.coli; EC3.
XX
OS Synthetic.

```
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..52
FT                   /note= "polymer repeat block sequence"
XX
PN   WO9524478-A1.
XX
PD   14-SEP-1995.
XX
PF   10-MAR-1995; 95WO-US02772.
XX
PR   11-MAR-1994; 94US-0212237.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J;
XX
DR   WPI; 1995-328270/42.
XX
PT   Protein polymer comprising alternating blocks of fibroin and elastin
PT   units - used to form a device e.g. a suture to keep separated
PT   viable tissue together

     Example 1: Page 22-24; 46pp; English.
XX
CC   The sequences represented by R80251-R80257 are polymers constructed of
CC   repeating blocks of fibroin-like (see R80249) and elastin-like (see
CC   R80250) units. This sequence contains 18 repeats of a block consisting
CC   of 8 elastin-like repeats followed by 2 fibroin-like repeats. The DNA
CC   sequences encoding these polymers were inserted into plasmids which were
CC   used to transform E.coli strain EC3. The polymers could then be
CC   isolated from the fermented strains by standard centrifugation
CC   techniques. The polymers are used to form a device (such as a suture,
CC   pin, thread, gel or film) to keep separated viable tissue together. By
CC   varying the the ratio of the two repetitive units, and by altering the
CC   lengths of the blocks of each of them, the tensile properties of the
CC   polymer can altered moderately. By reducing the number of repeating units
CC   of this sequence, or by increasing the number of units of the elastin
CC   like repeat, a faster rate of resorption can be achieved. Of the
CC   polymers, this sequence had the fastest resorption rate. The SELP4 (see
CC   R80255) and SELP5 (seeR80256) polymers showed no resorption after seven
CC   weeks. The other three polymers showed intermediate resorption. No data
CC   was given for SELP5 (see R80257).
XX
SQ   Sequence   936 AA;
```

```
Query Match             29.9%;  Score 1130;  DB 16;  Length 936;
Best Local Similarity   38.5%;  Pred. No. 1.4e-58;
Matches   377;  Conservative   46;  Mismatches 248;  Indels 308;  Gaps  62;

Qy        2 GVPG-AIPG-GVPG----GVPYPGAGLGALGGGALGPG------------------------------ 33
            ||||  ||  |:||    ||  ||: :||  |||
Db       10 gvpgvgvpgvgvpgvgvpgvgvpgvggagsgagssgagssvpgvgvpgvgvpgvgvpgvgvpgvgvp 69

Qy       34 --------------GKPIKPVPG-GLAGAGLAGAGLGGAPPAVPFPGALVPG------------ 58
                          | | |||:||| |:|:|:  ||:|
Db       70 gvpgvgvpgvgvpgvgvpgvgvpgvggagsgagssgagssvpgvgvpgvgvpgvgvpgvgvpgvgvg 129

Qy       69 ------GVADRAAAYKAAKAGAGLGGVPGVG------GLGVSA---GAVVP---QPG 107
                  || ||||:|| |:||:  |   |:||       ||:|    ||  || |
Db      130 VPgvgvgvYPgvgvPgvggagsgagssgagssvpgvgvpgvgvpgvgvpgvgvpgvgvPg 189

Qy      108 AGVK----------PGKVPGVGLPGV-TPGGVLPGARFPGVGVLPGVPT-GAGVKPFKA 153
               ||                  ||||| |    |:   ||||   ||||:  || ||
Db      190 vgvPgvggagsgagssgagssvpgvgvPgvgvPgvgvPgvgvp------pgvgvgvPgvgv---- 244

Qy      154 PGVGGAFAG------IPGVGGPGQ2GVPL-GYPIKAPKLPG----GYGLPYTTGSKLP 201
            ||||| ||       :|||  |||    | | |:|  :|  :|   :||:|
Db      245 pgvggagsgagssgagssvpgvgvpgvgvpgvgvpgvgvpgvgvpgvgvPgvgvPgvgga-- 302

Qy      202 YGYGPGGVAGAAGKAGIP-----------------------TGTGVGPQAAAAAAKAAAKPGA 242
            ||||| ||:|:||||           |    ||||   |||:
```

Qy (continued alignment blocks — database Db entries):

```
Db      303 -gagsgagssvpgvgvPgvgvPgvgvPgvgvpgvgvpgvgvpgvggagsgagssga 360
Qy      243 GAAGVLPGVG--GAGVP-GVPG-AIPGIG----GIAGVGTPAAAAAAARAAAAKYGAA 293
Db      361 g-agsvpgvgvPgvgvpgvgvPgvgvpgvgvpgvgvpgvggagss--------gag 413
Qy      294 AGLVP--GGPGFG-PGV----VGVPGAGVPGVPGVPGAGIP---------------- 326
Db      414 agsvpgvgvPgvgvPgvgvPgvgvPgvgvPgvgvPgvggagsgagssgagssvpgvg 473
Qy      327 --------VVPGAGIPGAAVPGVGVYSPEAAAKAAKAAKYGARPGVGVG--GIPT 370
Db      474 VPgvgvgvgYPgvgvPgvgvPgvgvPgvgvPgvggagsgagssvpgvgvpgvgvYPpg 533
Qy      371 YGVGAGGFPGRGV-GVG-----GIPGV----------------AGVP-SVG----GVPGVG- 404
Db      534 vgvPgvgvgvgYPgvgvPgvgvPgvgvPgvggagsgagssgagssvpgvgvpgvgvYPAPGVGVPgvgv 593
Qy      405 --GVPGVGISPEAQARAAAAKRAKYGVGTPAAAAAAKAAAKRAAAQFGLVPGVGVTAPGVGVAP 461
Db      594 pgvgvpgvgvp-----------------gvgvpgvgga-ggagsagssvpgvgv-pgvgvp--p 636
Qy      462 GVGVAPGVGLAPGVGVAPGVSVAPGVPGIPGGVAAAAKSHAKVAAKAQLRAAAGLG 521
Db      637 gvgv-pgvgv-pgvgv-pgvgv-pgvggagsgagssga-------------gagsvpg 679
Qy      522 AGIPGLGV-GVGVPGLGV-GAGVPGGV-GAGVPGF-GAGADEGVRRSLSPELRPGGDPSS 577
Db      680 vgvpgvgvpgvgvpgvgvpgvgvPgvgvPgvgvpgvgvpgvgvPgvgv--gvPgv 737
Qy      578 SQHLPSTP-----------------SSPKVPGALA---AAKAAKYGAAVPGV-LGGLGAL 616
Db      738 gvpgvgvpgvgvpgvgvpgvgvpgvgvpgvggagsgagssgagssvpgvgvPgvgvYPpgvp 797
Qy      617 G----GVGIPG-GVVGAG-PAAAAAAKAAAAKAAAQFGLVGAAGLG-------------- 655
Db      798 gvpgvgvpgvgvpgvgvPgvggagsgagssgagssvpgvgvpgvgvpgvgvPgvgvpgvg 857
Qy      656 --GLGVGGLGVPGVGGLGSIPPAAAAAKAAKYGAAGLG----GVLG-GAGQFPLGGVAARP 708
Db      858 VPgvgvYPgvgvPgvgvggagsgagssgagssvpgvgvpgvgvYPgvgvPgvgvPgvgvgvPgvgvp-p 916
Qy      709 GFGLSPIFPGGACLGKACG 727
Db      917 gvgvpgvp--ggagsgagssgag 933
```

```
RESULT 15
R80335
ID   R80335 standard; Protein; 877 AA.
XX
AC   R80335;
XX
DT   18-APR-1996 (first entry)
XX
DE   Protein polymeric adhesion substrate 1-G.
XX
KW   Pendent group; repeating unit; enzyme recognition site; sealant; fibrin;
KW   enzymatic cross-linking; biocompatible material; structural integrity;
KW   medical adhesive; wound closure; tissue repair; transglutaminase;
XX
OS   Synthetic.
XX
FH   Key        Location/Qualifiers
FT   Domain     37..53
FT              /note= "human fibrin gamma chain-derived domain"
FT   Domain     846..862
FT              /note= "human fibrin gamma chain-derived domain"
XX
PN   WO9523611-A1.
XX
PD   08-SEP-1995.
XX
```

PF   03-MAR-1995;   95WO-US02728.
XX
PR   03-MAR-1994;   94US-0205518.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J;
XX
DR   WPI; 1995-320413/41.
XX
PT   Protein polymers comprising repeating units and sequences - capable
PT   of enzyme-catalysed covalent bond;formation useful as a
PT   biocompatible material for wound closure and tissue repair
XX
PS   Example 6; Page 56; 138pp; English.
XX
CC   The amino acid sequence of the protein polymeric adhesion substrate
CC   (PPAS) 1-6. The protein comprises 11 repeats of the peptide sequence
CC   [(GVGVP)8(GAGAGS)4] inserted between 2 human fibrin gamma chain-derived
CC   peptide sequences. The protein can be used as a polymer substrate in a
CC   isopeptide cross-linking reaction;catalysed by the Factor XIII
CC   transglutaminase enzymatic activity. The polymers can be used in
CC   biological systems where in situ formation of a biocompatible material
CC   with structural integrity is required e.g. as medical adhesives and
CC   sealants or for wound closure or tissue repair.
XX
SQ   Sequence   877 AA;

Query Match   29.8%;   Score 1126.5;   DB 16;   Length 877;
Best Local Similarity   44.0%;   Pred. No. 2.1e-58;
Matches 347;   Conservative   59;   Mismatches 230;   Indels 153;   Gaps   57;

Qy   2   GVPG-AIPG-GVPG----GVFYPGAGLGAGLGGGALGPGGKPLKPVPGGLAGAGLGAGLG-   54
            |||| :|| |||| || || |||||| |: | || | :||| |:||:|
Db   84   gvpgvgvpgvgvpgvgvpgvgvpgvgvpgga-gaga------gagsgagsgagsgvgv-   134

Qy   55   ------AFPAVTFPGALVPG----GVADAAAAYKAAKAGALGGVPGVG-GLGVSAGA   101
                  || |: ||| ||| || | | | ||||||:| | | || ||
Db   135   pgvgvpgvgvpgvgvpgvgvpgvgvpgvpgsagsgagsgagsgagsgagsgagsgags   194

Qy   102   VVPQPGAGVKPGKVPGVGLPGV-YPGGVLPGARPPGVGVLPGVPTGAGV-KPKAPGVG-   157
                | | || || :|||:|| || ||:| || ||| |:| | :|
Db   195   gvgvpgvgvpgvgvpgvgvpgvgvpgvgv-pgagsgagsgagsgagsgagsg   253

Qy   158   ---GAFAGIPGVGFPGGPQPGVPLGYPIKAPXLPGGYGLPYTTGKLPYGVG-PG-GVAGA   212
                  |: ||||| | ||| | || || | ||:|||:||:| |: ||:
Db   254   agagsgvgvpgvgvpgvgv-gvp-------gvgvp------gvgvpgvvpg-   294

Qy   213   AGKAGYPTGTGVGPQAAAAAAAKAKAAKFGAGAAGV-LPGVG--GAGVPGVPGAIPGIG--   267
                |: :|| |:|| |: | | || | | | || | | | | ||:|
Db   295   vgvpgagsgagsgagsgagsgagsgagsgvgvYPgvgYPgvgYP-gvYPgvgvp   352

Qy   268   --GIAGVGTPAAAAAAAAAAKAAKYGAAAGLVPGGPGGPGVVGVPGAGVPGVGVPGAGI   325
              :: ||| || | | |:|: | | | || || |||||:||||| |:|| |
Db   353   gvgvpgvgvpgagsgags---gagags---sgtagsg-VgvPgvgvpgvgYPgvgvpgvgv   405

Qy   326   P---VVPGGIPGAAVPGVVSPEAAAAKAKAAKYGARPGGVGGIPYGVGAG---GEPG   380
                 |:|||||: ||:||||:: ::|:|:| | |:|| | || :| :
Db   406   PgvgvYPgvgYPgvgvpgvgvpgvgvpgvgags---gagsgagsgvgvPpg   456

Qy   381   EGVGVGGGIPGVAGVPSVGGVPGVG----GVPGVGISPEAQAAAAAKAAKYGVGTPAAAAA   436
                || :||| |||: | |:| | || || |:||||||| |:||| |||||||:|:
Db   457   vgvpgvgvpgv-gvpgv--gvpgvgvpgvgvpgvgvpgvpgv-----gagsgagags   508

Qy   437   KAAAKAAQFGLVPGGVAPGVGVAPGVGVLAPGVVAPGVGVAPGVGVTAPGIGPFG   496
                |:|| | | | || |:||| |:|| | ||||||| |||:| |||:|:||:|| ||
Db   509   gagags------gvgvgv-pgvgv-pgvgvgv-pgvgv-pgvgvp   554

Qy   497   GVAAAAKSAAKVAAKAQLRAAAGLSAGIPGLGV------GVGVPGLGV-GAGVPGLGV-G   548
                || ||: | | | ||||:| | || ::|| | ||||:||| ||||:||: | |
Db   555   gagsgagsgagsgagsgagsscagagsgvgvpgvgvpgvgvpgvgvpgvgvpgvgvYPgvgvpg   614

Qy   549   AGVPGFGAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPG   608
                |||| ||: |||:: | : | | | :: | | | | |:||||||:|||||
Db   615   vgvpgagsgagsga---------gsagsgagagsgvgvpg-vgvpgvgvpgvgvpgvgvpg   663

Qy   609   V-LGGLGALG----GVGIPGGVVGAGPAAAAAKAAAAKAAKAAQFGL----VGRAGLG--GL   657
                | | || | | | || | | | | || | | || || |:| |||:|:| ||
Db   664   vgvpgvgvpgvgvpgvgvpgvgvpgagsgagsgagsgvgvpgvgvpgvgvpgvgvpgv   723

Qy   658   GVGGLGVPGVG--GLG----GIPPAAAAKAAKYGA-AGLGGVLGGASQFPLGGVAARPGF   710
                |:|:||||||: |:| : | ||||||:|:|| :| |: || | | |:| || |
Db   724   gvpgvgvpgvgvpgvgvpgvgvpgagsgagsgagsgagsgvgvpgv-pgv-pgv   782

Qy   711   GLSPI-FPG   718
                |: : ||
Db   783   gvpgvgvpg   791

Search completed: April 24, 2001, 16:38:14
Job time: 414 sec

```
DR   EMBL; U52292; AAB17544.1; -.
DR   PIR; A30524; A30524.
DR   HSSP; P04002; 1WFA.
DR   MIM; 130160; -.
DR   MIM; 194050; -.
KW   Structural protein; Connective tissue; Repeat; Signal.
FT   SIGNAL        1     26
FT   CHAIN        27    730       ELASTIN.
SQ   SEQUENCE   730 AA;  63260 MW;  AB06D15BA567A3A6 CRC64;


Query Match             94.5%;  Score 3575.5;  DB 1;  Length 730;
Best Local Similarity   94.7%;  Pred. No. 2.7e-134;
Matches  698;  Conservative  0;  Mismatches  0;  Indels 39;  Gaps  2;

Qy    1 GGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPSGLAGAGLGAGLGAFPAVT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27 GGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPSGLAGAGLGAGLGAFPAVT 86

Qy   61 FPGALVPGGVADAAAAYKAAAKAGAGLGGVPGVGLGGSAGAVPGPGAGVKPGKVPGVGL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87 FPGALVPGGVADAAAAYKAAAKAGAGLGGVPGVGLGGSAGAVPGPGAGVKPGKVPGVGL 146

Qy  121 PGVTPGGVLPGARPGVLPGVPLPGVPTGAGVKPKPKAPGVPGGAFPAGIPGVGPGVPLGY 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  147 PGVTPGGVLPGARPGVCVLPGVPLPGVPTGAGVKPKPKAPGVPGGAFPAGIPGVGPGVPLGY 206

Qy  181 PIKAPKLPGGVGLPYTTGKLPYGYGPGGVAGAGKAGYPTGTGVGPQAAAAAAKAKAKF 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  207 PIKAPKLPGGVGLPYTTGKLPYGYGPGGVAGAGKAGYPTGTGVGPQAAAAAAKAKAKF 266

Qy  241 GAGAAGVLPGVGGAGVPGVPGVPGAIPGIGGIAGVGTPAAAAAAAAKAKYGAAAGLVPGG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  267 GAGAAGVLPGVGGAGVPGVPGVPGAIPGIGGIAGVGTPAAAAAAAAKAKYGAAAGLVPGG 326

Qy  301 PGPGPGVVGVPGAGVPVVPGAGIPGAAVPGVVSPEAAAKAAAAKAAKYGAR 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  327 PGPGPGVVGVPGAGVPVVPGAGIPGAAVPGVVSPEAAAKAAAAKAAKYGAR 386

Qy  361 PGVGVGGIPGVGVGAGGPGFGVGVGGIPGVAGVPSVGVGGVPGVGGVPGVGGIPGVSPEAQAAAA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  387 PGVGVGGIPGVGVGAGGPGFGVGVGGIPGVAGVPSVGVGGVPGVGGVPGVGGIPGVSPEAQAAAA 446

Qy  421 AKAKAYGVPGAAAAAAKAAAQF------GLVPGVGVAPGVGVAPGVGVAPGVGLAPG 474
          |||||||||||||||||||          |||||||||||||||||||||||||||
Db  447 AKAKAYGVPGAAAAAAKAAAQFALLNLAGLVPGVGVAPGVGVAPGVGVAPGVGLAPG 505

Qy  475 VGVAPGVGVAPGVGVAPGIGPGVGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGVP 534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  507 VGVAPGVGVAPGVGVAPGIGPGVGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGVP 566

Qy  535 GLGVGAGVPGLGVGAGVPGVGAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGAL 594
          |||||||||||||||||||||
Db  567 GLGVGAGVPGLGVGAGVPGFGA-------------------------------YPGAL 593

Qy  595 AAAKAAKYGAAVPGVGVLGGIGAGLGGVGTPGGVGAGPAAAAAAKAAAAKAAQFGLVGAAGL 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  594 AAAKAAKYGAAVPGVGVLGGIGAGLGGVGTPGGVGVSAGPAAAAAAKAAAAKAAQFGLVGAAGL 653

Qy  655 GGLGVGGLGVPGVGVGGLGGIGTPPAAAAAKAAKYGAAAGLGGVLGGAGQPPLGGVAARPGFGLSP 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  654 GGLGVGGLGVPGVGVGGLGGLGALGGVGTPPAAAAAKAAKYGAAAGLGGVLGGAGQPPLGGVAARPGFGLSP 713

Qy  715 IFPGGACLGKACGRKRK 731
          |||||||||||||||||
Db  714 IFPGGACLGKACGRKRK 730
```
```
RESULT 2
ELS_BOVIN
ID   ELS_BOVIN     STANDARD;     PRT;    747 AA.
AC   P04985; P04987; Q29421;
DT   13-AUG-1987 (Rel. 05, Created)
```

```
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN).
GN   ELN.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87194772; PubMed=3032943;
RA   Raju K., Anwar R.A.;
RT   "Primary structures of bovine elastin a, b, and c deduced from the
RT   sequences of cDNA clones.";
RL   J. Biol. Chem. 262:5755-5762(1987).
RN   [2]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   TISSUE=Nuchal ligament;
RX   MEDLINE=89274159; PubMed=2543440;
RA   Yeh H., Anderson N., Ornstein-Goldstein N., Bashir M.M.,
RA   Rosenbloom J.C., Abrams W.R., Indik Z., Yoon K., Parks W., Mecham R.,
RA   Rosenbloom J.;
RT   "Structure of the bovine elastin gene and S1 nuclease analysis of
RT   alternative splicing of elastin mRNA in the bovine nuchal ligament.";
RL   Biochemistry 28:2365-2370(1989).
RN   [3]
RP   SEQUENCE OF 1-27 FROM N.A.
RX   MEDLINE=91234332; PubMed=2031719;
RA   Manohar A., Shi W., Anwar R.A.;
RT   "Partial characterization of bovine elastin gene; comparison with the
RT   gene for human elastin.";
RL   Biochem. Cell Biol. 69:185-192(1991).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A (SHOWN HERE), B AND C; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
```

```
DR   EMBL; J02717; AAA30503.1; -.
DR   EMBL; K03505; AAA30505.1; -.
DR   EMBL; K03506; AAA30506.1; -.
DR   EMBL; J02855; AAA30776.1; -.
DR   EMBL; M58652; AAA03519.1; -.
DR   PIR; A26728; A26728.
DR   PIR; B26728; B26728.
DR   PIR; C26728; C26728.
DR   HSSP; P04002; 1WFA.
KW   Structural protein; Connective tissue; Repeat; Signal;
KW   Alternative splicing.
FT   SIGNAL        1     26
FT   CHAIN        27    747       ELASTIN.
FT   MOD_RES     105    105       OXIDATIVE DEAMINATION.
FT   MOD_RES     109    109       OXIDATIVE DEAMINATION.
FT   MOD_RES     252    252       OXIDATIVE DEAMINATION.
FT   MOD_RES     271    271       OXIDATIVE DEAMINATION.
FT   MOD_RES     275    275       OXIDATIVE DEAMINATION.
FT   MOD_RES     324    324       OXIDATIVE DEAMINATION.
FT   MOD_RES     327    327       OXIDATIVE DEAMINATION.
FT   MOD_RES     400    400       OXIDATIVE DEAMINATION.
FT   MOD_RES     404    404       OXIDATIVE DEAMINATION.
FT   MOD_RES     407    407       OXIDATIVE DEAMINATION.
```

```
                    GenCore version 4.5
             Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      April 24, 2001. 16:35:14 ; Search time 44.88 seconds
                                        (without alignments)
                                        557.949 Million cell updates/sec

Title:
Perfect score:     3785
Sequence:          1 GGVPGAIPGGGVPGGVFYPGA.......LSPIFPGGACLGKACGKKKk 731

Scoring table:     BLOSUM62
                   Gapop 10.0 , Gapext 0.5

Searched:          93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters:       93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  SwissProt_39:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3575.5	94.5	730	1	ELS_HUMAN	P15502 homo sapien
2	2479	65.5	747	1	ELS_BOVIN	P04985 bos taurus
3	2387	63.1	860	1	ELS_MOUSE	P54320 mus musculu
4	2311	61.1	864	1	ELS_RAT	Q99372 rattus norv
5	1542.5	40.8	750	1	ELS_CHICK	P07916 gallus gall
6	719	19.0	5263	1	PEOH_BOMMO	P05790 bombyx mori
7	675.5	17.8	1758	1	CA24_CAEEL	P17140 caenorhabdi
8	663	17.5	1901	1	Y208_MYCTU	O53553 mycobacteri
9	644	17.0	747	1	SPD1_NEPCL	P19837 nephila cla
10	609.5	16.1	1763	1	CA14_HUMAN	P27393 ascaris suu
11	609.5	16.1	1685	1	CA54_HUMAN	P02462 homo sapien
12	599.5	15.8	1685	1	CA54_HUMAN	Q29400 homo sapien
13	589.5	15.6	1669	1	CA14_MOUSE	P02463 mus musculu
14	580	15.3	778	1	YQ34_MYCTU	P71933 mycobacteri
15	580	15.3	1464	1	CA11_HUMAN	P02452 homo sapien
16	578	15.3	1460	1	CA11_CANFA	Q9XS17 canis famil
17	577.5	15.3	1453	1	CA11_CHICK	P02457 gallus gall
18	575.5	15.2	1456	1	CA13_HUMAN	P02460 homo sapien
19	572	15.1	1364	1	CA21_BOVIN	P02465 bos taurus
20	570.5	15.1	1356	1	CA21_CANFA	O46392 canis famil
21	570.5	15.1	1678	1	CA64_HUMAN	Q14031 homo sapien
22	570	15.1	1049	1	CA13_BOVIN	P04258 bos taurus
23	563	14.9	1418	1	CA12_HUMAN	P02458 homo sapien
24	560.5	14.8	1453	1	CA11_MOUSE	P11087 mus musculu
25	560	14.8	1464	1	CA13_MOUSE	P08121 mus musculu
26	559	14.8	1373	1	CA21_MOUSE	Q01149 mus musculu
27	556	14.7	1459	1	CA12_MOUSE	P28481 mus musculu
28	553.5	14.6	1366	1	CA21_HUMAN	P08123 homo sapien
29	553	14.6	1712	1	CA24_HUMAN	P08572 homo sapien
30	551	14.6	957	1	Y278_MYCTU	P56877 mycobacteri
31	543.5	14.4	1362	1	CA21_CHICK	P02467 gallus gall
32	543	14.4	914	1	WA22_MYCTU	O06794 mycobacteri
33	539.5	14.3	1758	1	CA14_CAEEL	P17139 caenorhabdi
34		539	14.2	627	1	SPD2_NEPCL	P46804 nephila cla
35		533.5	14.1	1496	1	CA25_HUMAN	P05997 homo sapien
35		531	14.0	1707	1	CA24_MOUSE	P08122 mus musculu
37		530.5	14.0	801	1	Y747_MYCTU	O53810 mycobacteri
38		526.5	13.9	779	1	CA11_BOVIN	P02453 bos taurus
39		525.5	13.9	1027	1	CAFF_RIFPA	P30754 riftia pach
40		514.5	13.6	1603	1	CA1F_HUMAN	Q07092 homo sapien
41		513	13.6	1690	1	CA44_HUMAN	P53420 homo sapien
42		505	13.3	1806	1	CA1B_HUMAN	P12107 homo sapien
43		503.5	13.3	1736	1	CA2B_HUMAN	P13942 homo sapien
44		502	13.3	1804	1	CA1B_MOUSE	Q61245 mus musculu
45		501	13.2	1670	1	CA34_HUMAN	Q01955 homo sapien

ALIGNMENTS

```
RESULT 1
ELS_HUMAN
ID   ELS_HUMAN      STANDARD;      PRT;     730 AA.
AC   P15502;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN).
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin fibroblast;
RX   MEDLINE=89009960; PubMed=3171221;
RA   Fazio M.J., Olsen D.R., Kauh E.A.,  Baldwin C.T., Indik Z.,
RA   Ornstein-Goldstein N., Yeh H., Rosenbloom J., Ditto J.;
RT   "Cloning of full length elastin cDNAs from a human skin fibroblast
RT   recombinant cDNA library; further elucidation of alternative splicing
RT   utilizing exon-specific oligonucleotides.";
RL   J. Invest. Dermatol. 91:458-464(1988).
RN   [2]
RP   SEQUENCE OF 603-730 FROM N.A.
RC   TISSUE=Placenta, and Hippocampus;
RX   MEDLINE=96291399; PubMed=8689688;
RA   Frangiskakis J.M., Ewart A.K., Morris C.A., Mervis C.B.,
RA   Bertrand J., Robinson B.F., Klein B.P., Ensing G.J., Everett L.A.,
RA   Green E.D., Proeschel C., Gutowski N.J., Noble M., Atkinson D.L.,
RA   Odelberg S.J., Keating M.T.;
RT   "LIM-kinase1 hemizygosity implicated in impaired visuospatial
RT   constructive cognition.";
RL   Cell 86:59-69(1996).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   -!- DISEASE: DELETED IN WILLIAMS-BEUREN SYNDROME (WBS), A
CC       DEVELOPMENTAL DISORDER CAUSED BY HAPLOINSUFFICIENCY FOR GENES IN A
CC       2-CM REGION OF CHROMOSOME BAND 7Q11.23. HEMIZYGOUS DELETION MAY
CC       CONTRIBUTE TO CERTAIN DEFECTS SUCH AS HYPERCALCEMIA AND GROWTH
CC       DELAY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on  its
CC   use  by  non-profit institutions as long as its content is  it in  no  way
CC   modified and this statement is not removed. Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X15603; CAA33627.1; -.
DR   EMBL; M36860; AAA52382.1; -.
```

```
              GenCore version 4.5
         Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      April 24, 2001, 16:31:20 ; Search time 62.39 Seconds
                                         (without alignments)
                                         225.086 Million cell updates/sec

Title:          US-09-340-736-1
Perfect score:  3785
Sequence:       1 GGVPGAIPGGVPGGVFYPGA.........LSPIFPGGACLGKACGRKKK 731

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

?ched:          185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters:    185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued_Patents_AA:*
       1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
       2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
       3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
       4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
       5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
       6: /cgn2_6/ptodata/2/iaa/Backfiles1.pep:*

                       SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Result No.	Score	Query Match	Length	DB	ID	Description
1	3785	100.0	731	2	US-08-911-364-1	Sequence 15, Appl
2	3747.5	99.0	792	2	US-08-678-039A-40	Sequence 40, Appl
3	1159.5	30.6	988	5	US-08-212-237-5	Sequence 5, Appl
4	1159.5	30.6	988	5	PCT-US95-02772-5	Sequence 5, Appl
5	1159	30.6	832	4	US-08-212-237-4	Sequence 4, Appl
6	1159	30.6	832	4	PCT-US95-02772-4	Sequence 4, Appl
7	1130	29.9	936	3	US-08-212-237-3	Sequence 3, Appl
8	1130	29.9	936	3	PCT-US95-02772-3	Sequence 3, Appl
9	1126.5	29.8	877	1	US-08-397-633A-54	Sequence 54, Appl
10	1123	29.7	884	1	US-08-397-633A-68	Sequence 68, Appl
11	1123	29.7	884	2	US-08-435-641-15	Sequence 15, Appl
12	1123	29.7	884	2	US-03-642-246-15	Sequence 96, Appl
13	1123	29.7	884	5	US-03-642-246-15	Sequence 15, Appl
14	1123	29.7	884	5	PCT-US96-06229-15	Sequence 15, Appl
15	1122	29.6	1002	5	US-08-707-237A-103	Sequence 103, App
16	1122	29.6	1002	3	US-08-642-246-25	Sequence 25, Appl
17	1122	29.6	1002	5	PCT-US96-06229-25	Sequence 25, Appl
18	1118	29.5	1413	1	US-08-175-155-39	Sequence 39, Appl
19	1118	29.5	1413	1	US-08-707-237A-45	Sequence 45, Appl
20	1118	29.5	1464	1	US-08-477-509B-74	Sequence 74, Appl
21	1118	29.5	1464	3	US-08-482-085B-74	Sequence 74, Appl
22	1116	29.5	1056	5	US-08-212-237-6	Sequence 6, Appl
23	1116	29.5	1056	5	PCT-US95-02772-6	Sequence 47, Appl
24	1107	29.2	2257	1	US-08-175-155-47	Sequence 47, Appl
25	1107	29.2	2257	1	US-08-477-509B-82	Sequence 82, Appl
26	1107	29.2	2257	3	US-08-707-237A-53	Sequence 53, Appl
27	1107	29.2	2257	3	US-08-482-085B-82	Sequence 82, Appl
28	1100	29.1	936	2	US-08-707-237A-108	Sequence 108, App
29	1100	29.1	936	3	US-08-642-246-30	Sequence 30, Appl
30	1100	29.1	936	5	PCT-US96-06229-30	Sequence 30, Appl
31	1095	28.9	972	5	US-08-212-237-7	Sequence 7, Appl
32	1095	28.9	972	5	PCT-US95-02772-7	Sequence 7, Appl
33	1070.5	28.3	2055	1	US-08-175-155-46	Sequence 46, Appl
34	1070.5	28.3	2055	1	US-08-477-509B-81	Sequence 81, Appl
35	1070.5	28.3	2055	1	US-08-707-237A-52	Sequence 52, Appl
36	1070.5	28.3	2055	3	US-08-482-085B-81	Sequence 81, Appl
37	1069.5	28.3	1024	1	US-08-212-237-8	Sequence 8, Appl
38	1069.5	28.3	1024	5	PCT-US95-02772-8	Sequence 8, Appl
39	1049	27.7	966	3	US-08-642-246-34	Sequence 34, Appl
40	1049	27.7	966	5	PCT-US96-06229-34	Sequence 34, Appl
41	975.5	25.8	859	1	US-08-175-155-42	Sequence 42, Appl
42	975.5	25.8	859	1	US-08-477-509B-77	Sequence 77, Appl
43	975.5	25.8	859	1	US-08-707-237A-48	Sequence 48, Appl
44	975.5	25.8	859	3	US-08-482-085B-77	Sequence 77, Appl
45	975.5	25.8	1250	2	US-08-735-692-28	Sequence 28, Appl

ALIGNMENTS

```
RESULT
US-08-911-364-1
;-Sequence-1-Application US/08911364
; Patent No. 5969106
; GENERAL INFORMATION:
;  APPLICANT: ROTHSTEIN, Aser
;  APPLICANT: ROTHSTEIN, Steven J.
;  APPLICANT: KEELY, Fred W.
;  TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
;  TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
;  NUMBER OF SEQUENCES: 8
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FOLEY & LARDNER
;   STREET: 3000 K Street, N.W.
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20007-5109
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/911,364
;   FILING DATE: 07-AUG-1997
;   CLASSIFICATION: 530
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/023,552
;   FILING DATE: 07-AUG-1996
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Bent, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 041082/0104
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 672-5300
;   TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 1:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 731 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;  MOLECULE TYPE: peptide
US-08-911-364-1

Query Match             100.0%;  Score 3785;  DB 2;  Length 731;
Best Local Similarity   100.0%;  Pred. No. 5e-246;
Matches  731;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy      1 GGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT  60

Qy     61 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL  120

Qy    121 PGVYPGGVLPGARPGVGVLPGVPTGAGVKPKPAPGVGGAFAGIPGVGPFGGPGPGVPLGY  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 PGVYPGGVLPGARPGVGVLPGVPTGAGVKPKPAPGVGGAFAGIPGVGPFGGPGPGVPLGY  180

Qy    181 PIKAPKLPGGYGLPYTTGKLPYTTSKLPYGTPGGVAGAASKAGYPTGTGVGPQAAAAKAAAAKF  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 PIKAPKLPGGYGLPYTTGKLPYTTSKLPYGTPGGVAGAASKAGYPTGTGVGPQAAAAKAAAAKF  240

Qy    241 GAGAAGVLPGVGGAGVPGVGVPGVGPGVGVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGAR  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGAAGVLPGVGGAGVPGVGVPGVGPGVGVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGAR  300

Qy    301 PGFGPGVGVPGAGVPGVGPGVGPGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGAR  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 PGFGPGVGVPGAGVPGVGPGVGPGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGAR  360

Qy    361 PGVGVGGIPYTGVGAGGFPGFGVGVGGIPGVAGYPSVGGVPGVGGYPGVGGVGGIPGVAGYP  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 PGVGVGGIPYTGVGAGGFPGFGVGVGGIPGVAGYPSVGGVPGVGGYPGVGGVGGIPGVAGYP  420

Qy    421 AKAAKYGVSTPAAAAAKAAAKAAQFGLVPGVGVGVAPGVVAPGVGVARPGVGVGVAPG  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AKAAKYGVSTPAAAAAKAAAKAAQFGLVPGVGVGVAPGVVAPGVGVARPGVGVGVAPG  480

Qy    481 VGVAPGVGVAPGIGPGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGPGLGVGA  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 VGVAPGVGVAPGIGPGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGPGLGVGA  540

Qy    541 GVPGLGVGAGVPGFGAGADEGVRRKSLSPELREGDFSSSQHLPSTPSSPRVPGALAAAKAA  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GVPGLGVGAGVPGFGAGADEGVRRSLSPELREGDFSSSQHLPSTPSSPRVPGALAAAKAA  600

Qy    601 KYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAKAAAKAAQFGLVGRAGLGGLGVG  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 KYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAKAAAKAAKAAQFGLVGRAGLGGLGVG  660

Qy    661 GLGVPGVGGLGGLGGIPPAAAAKYAAKYGAAGLGGVLGGAGQFPLGGVAARPGFGLSFIPPGGA  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GLGVPGVGGLGGLGGIPPAAAAKYAAKYGAAGLGGVLGGAGQFPLGGVAARPGFGLSFIPPGGA  720

Qy    721 CLGXACGRKRK 731
          |||||||||||
Db    721 CLGXACGRKRK 731
```

```
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/678,039A
    FILING DATE: 10-JUL-1996
    CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
    NAME: Saxe, Stephen A.
    REGISTRATION NUMBER: 38,609
    REFERENCE/DOCKET NUMBER: 2323-120A
TELECOMMUNICATION INFORMATION:
    TELEPHONE: 202-624-1589
    TELEFAX: 202-783-6031
INFORMATION FOR SEQ ID NO: 40:
    SEQUENCE CHARACTERISTICS:
    LENGTH: 792 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
    MOLECULE TYPE: protein
US-08-678-039A-40
```

```
Query Match        99.0%;  Score 3747.5;  DB 2;  Length 792;
Best Local Similarity  95.4%;  Pred. No. 2.1e-243;
Matches 731; Conservative  0; Mismatches  0;  Indels  35;  Gaps   2;
```

```
Qy      1 GGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     27 GGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT  86

Qy     61 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     87 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL  146

Qy    121 PGVYPGGVLPGARPGVGVLPGVPTGAGVKPKPAPGVGGAFAGIPGVGPFGGPGPGVPLGY  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    147 PGVYPGGVLPGARPGVGVLPGVPTGAGVKPKPAPGVGGAFAGIPGVGPFGGPGPGVPLGY  206

Qy    181 PIKAPKLPGGYGLPYTTSKLPYGTPGGVAGAGYPTGTGVGPQAAAAAKAAAKF  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    207 PIKAPKLPGGYGLPYTTSKLPYGTPGGVAGAGYPTGTGVGPQAAAAAKAAAKF  266

Qy    241 GAGAAGVLPGVGGAGVPGVGVPGVGPGVGVPGAGIPGGIAGVGTPAAAAAAKAAAKYGAAGLVPGG  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    267 GAGAAGVLPGVGGAGVPGVGVPGVGPGVGVPGAGIPGGIAGVGTPAAAAAAKAAAKYGAAGLVPGG  326

Qy    301 PGPGPGVGVPGAGVPGVGVGGAGGPGFGVGVGGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGAR  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    327 PGFGPGVGVPGAGVPGVGVGGAGGPGFGVGVGGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGAR  386

Qy    361 PGVGVGGIPTGVGAGGPGFGVGVGGIPGVAGYPSVGGIPGVAGYPGVGGVPGVGISPEAQAAAAA  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    387 PGVGVGGIPTGVGAGGPGFGVGVGGIPGVAGYPSVGGIPGVAGYPGVGGVPGVGISPEAQAAAAA  446

Qy    421 AKAAKY----------------------------------GVGTPAAAAKAAAKAAQF--------  445
          ||||||                                  |||||||||||||||
Db    447 AKAAKYGAAGAGVLGGLGGLVPGPQAAVPGVPGVPGVGVGTPAAAAKAAAKAAQPALLNLA  506

Qy    446 GLVPGVGVAPGVGVAPGVGVGVAPGVGVLAPGVGVGVARPGVGVAPGIGPGVVAAAAKSA  505
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    507 GLVPGVGVAPGVGVAPGVGVGVAPGVGVLAPGVGVGVARPGVGVAPGIGPGVVAAAAKSA  566

Qy    506 AKVAAKAQLRAAAGLGAGIPGLGVGVGPGFGAGADEGVRRS  565
          |||||||||||||||||||||||||||||||||||||||||
Db    567 AKVAAKAQLRAAAGLGAGIPGLGVGVGPGFGAGADEGVRRS  626

Qy    566 LSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGVLGGLGALGGVGIPGGV  625
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    627 LSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGVLGGLGALGGVGIPGGV  686
```

```
Qy   626  VGAGPAAAAAKAAKAAKAAQFGLVGAAGLGGLGVGGLGVGGLGVGGLGVGTPGVGGLGGIPPAAAAKAAKYG  685
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   687  VGAGPAAAAAKAAKAAKAAQFGLVGAAGLGGLGVGGLGVGGLGVGGLGVGTPGVGGLGGIPPAAAAKAAKYG  746

Qy   686  AAGLGGVLGGAGQFPLGGVAARPGFGLSPIFPGGACLGKACGRKR  731
          |||||||||||||||||||||||||||||||||||||||||||||
Db   747  AAGLGGVLGGAGQFPLGGVAARPGFGLSPIFPGGACLGKACGRKR  792
```

RESULT 3
US-08-212-237-5
; Sequence 5, Application US/08212237
; Patent No. 5605019
; GENERAL INFORMATION:
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Synthetic Proteins As Implantables
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,237
; FILING DATE: 11-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-58847/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 988 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-212-237-5

Query Match   30.6%; Score 1159.5; DB 1; Length 988;
Best Local Similarity 40.5%; Pred. No. 2.6e-70;
Matches 353; Conservative 54; Mismatches 279; Indels 185; Gaps 51;

```
Qy     2  GVPG-AIPG-GVPG------GVFYPGAGLGALGGGALGPG-------GKPLKPVPGGLGAGAG   48
          ||||  ||| ||||       ||:  ||:|    ||        |   ||||
Db    10  GVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGAGSSGAGAGSSGAGAGSGAGAGSSGAGAG   69

Qy    49  LGAGLGAPPAVTFPGALIVPGGVADAAAAYKAAKAGAGLG--GVPGVGGLGVSAGAVPQ-  105
          |||     |||: ||  :  |                  |:::  ||||||:  :  :||
Db    70  SGAGAGSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGSGAGAGSSGAGAGSGAGAGSSG  129

Qy   106  ----------PGAGVKPGKTPGVGLPGV-YPGGVLPGARFPGVGVLPGVPT-GAGV  149
                     |||   | | ||||:|   :  || :   | | ||||
Db   130  AGAGSSGAGAGSGAGAGSSGAGAGSVPGVGVPGVGVPGVGVPGVGV-PGVGGSGV  188

Qy   150  KPKAPGVGGAFAG-----------------------------IPGVGPFGGPGPQPGVPL  178
          |||| ||||                                 :||   ||:|| |
Db   189  ----PGVGGAGAGSSGAGAGSGAGAGSSGAGAGSGAGAGSVPGVGV---PGVGVP-  240

Qy   179  GYPIKAPKLPG----GVGLPYTTGKLPYGYGPGGVAGAAGKAGIPTGTGVGPQAAAAAAA  234
          |  : |||:||      || |   :|  | | |     |    :|  :|:|||||||
Db   241  GVGVPGVGVGVPGVGVP-------GVGVPGGAGAGSSGAGAGSSGAGAGS  292

Qy   235  KAAAKFGAGAAGVLPGVG--GAGVP--GVPG-AIPGIG----GIAGVGTPAAAAARAAA  285
          |  :  ||    |:|||     |||  ||||    :::     ||:||:|
Db   293  GAGAGSSGAG-AGSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGAGSG  351

Qy   286  KAAKYGAAAGLVPG------GPGPGVGVPGGAGVPGVGVPGCAGIP--VVPGAGIPGAAV  338
          |  | | :  :||        |||||:||| | ||||||||     :|  ||  : |:
Db   352  AGAGSGAGAGSSGAGAGSGAGAGSGAGAGSVPGVGVPGVGVPGVGVPGVGVPGVGV  411

Qy   339  PGTVSPRAAAKAAAKAAKYGARPGVGVGGIPTYGVGAGGFPGFGVGGIPGV----AGV  394
          ||:|     | |  |  :|  | | ||:|   |  | ::     :|  ||||    |||
Db   412  PGVGTPGVGGAGAGSGAGAGSSGA----GAGSGAGAGSGAGSVPGVGVPGVGV  467

Qy   395  PSVG----GVPGVG----GVPGVGVPGVGVPGVGISPEAQRAAAAAKAAKYGVGTPAAAAKA----AAK  441
          | ||    |||| |    |||||||||||||||:  | ||                           |||
Db   468  PGVGVPGVGVPGVPGVGVPGVGV-PGVGGAGAGSGAGAGSGAGAGSSGAGAGS  526

Qy   442  AAGFGLVPGVGVAPGVGVAPGVGLAPGVGVAPGVGIGPGGVRAA  501
          |||  :||||||:||||| ||||  ||||| ||||:| | :|||
Db   527  GAGAGSVPGVGV-PGVGV-PGVGV-PGVGV-PGVGGAGAGSG  579

Qy   502  AKSAKVAAKADLRAAAGLGA---------GIPGLGV-GVGVPGLGV-GAGVPGLGV  547
          |  :|   | |  :|||| |          ||||| |   |||| | | |||| |
Db   580  AGAGSGAGAGSGAGAGSSGAGAGSVPGVGVPGVGVPGVGVPGVGVPGVGV  639

Qy   548  -GAGVPGF-GAGADGVRRSLSPELRBG-DPSSSQHLPSTPSSPRVPGALAAAKAAKYGA  604
           ||||||  | | :|        |::   |      |     || ||| |
Db   640  PGVGVPGVGGAGAGSGAGAGSSGAGAGSGAGAGSVPG-VGVPGVGVPGV  698

Qy   605  AVPGV-LGGLGALG---GVGIPG-GVVGAGPAAAAAAKAAAKAAAKAAQFGL----------  648
          ||||  :|| | |     || || |:  |  |||||||| | | |  :
Db   699  GVPGVGVPGVGVPGVGVPGVGVPGVGGAGAGSGAGAGSSGAGAGSGAGA  758

Qy   649  ----VGAAGLG--GLGAVGGLGVPGVG--GLG----GIPPAAAKAAKYGAAGLG-----  690
              :| | |    | | ||:||||||   ||     |:|||| | | :  |||
Db   759  GSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGAGSGAGAGSSGAGAGS  818

Qy   691  -GVLGGAGQPPLGGVAARPGFGLSPI-FPG  718
           || |||   |||||||  ||||| |: |||
Db   819  GAGAGSSGAGAGSGAGAGSVPGVGVPGVGVPG  849
```

RESULT 4
PCT-US95-02772-5
; Sequence 5, Application PC/TUS9502772
; GENERAL INFORMATION:
; APPLICANT: Protein Polymer Technologies, Inc.
; TITLE OF INVENTION: Synthetic Proteins As Implantables
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02772
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: FP-58847-1-PC/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 5:

```
;  SEQUENCE CHARACTERISTICS:
;    LENGTH:  988 amino acids
;    TYPE:  amino acid
;    STRANDEDNESS:  single
;    TOPOLOGY:  linear
;  MOLECULE TYPE:  protein
PCT-US95-02772-5

Query Match           30.6%;  Score 1159.5;  DB 5;  Length 988;
Best Local Similarity 40.5%;  Pred. No. 2.6e-70;
Matches 353;  Conservative  54;  Mismatches 279;  Indels 185;  Gaps  51;

Qy     2  GVPG-AIPG-GVPG----GVFYPGAGLGALGGGALGPG--------GKPLKPVPCGLAGNG  48
          ||||  :|| ||      ||  | :|| | ||  ||        ::|
Db    10  GVPGGVGVPGVPGVGVPGVGVPGVGVPGVGVPGVGGAGSGAGSSGAGSGAGSGAGSGAGAG  69

Qy    49  LGAGLGAPPAVTFPGAILVFGGVADAAAAYKAKAGAGLG--GVPGVGGLGVSAGAVVPG-  105
          ||| |: |  | | :||||   || | |||| |   :: | | |
Db    70  SGAGAGSSGAGSGAGSGAGSVPGVGVPGVGVPGVGVPGVGGAGSGAGSGAGSGAGSG    129

Qy   106  --------------PGAGYKPGKVPGVGYLPGV-YPGGVLPGAARPGVGVLPGVPT-GAGV  149
                        |||||||:|:|||    || ||||| |  ||||| ||||
Db   130  AGAGSGAGSGAGSGAGSGAGSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGV       188

Qy   150  KPKAPGVGGAFAG--------------------------------IPGVGPFGGPGQPGVPL  178
          ||||||||                                   :||| 
Db   189  ----PGVGGAGSGAGSGAGSGAGSGAGSGAGSGAGSVPGVGV---PGVGVP-      240

Qy   179  GYPIKAPKLPG----GVGLPYTTGKLPYGZYPGGVYAGRAGKAGTPTGTGVGPQAAAAAAAA  234
          ||||: | :       ||| ||| |       |:|:|| | :|:| || |:|||||||||||
Db   241  GVGVPGVGVPGVGVP--------GVGVPGVGGAGSGAGSGAGSGAGSGAGS           292

Qy   235  KAAAKFPGAGAAGVLPGVG---GVPG-AIPGIG----GIAGVGTPAAAAAAAAAA      285
          |:| | | :||| || ||   :||| ||||||     ||:| | ||||||||||||
Db   293  GAGAGSGAG-AGSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGSGSG     351

Qy   286  KAAKYGAAAGUVPG-----GPGFGPGVVGVPGAGYPGVGVPGAGIP--VVPGAGIPGAAV   338
          |:| | |  |         |||| | || :| ||||:|| |     | :||  |:|
Db   352  AGAGSGAGSGAGSGAGSGAGSGAGSGAGSVPGVGVPGVGVPGVGVPGVGVPGVGV       411

Qy   339  PGVVSPEAAAKAAAKAAKAAKYGARPGVGVGGIPTIYGVGAGGPPGPGVGVGGIPGV----AGV  394
          |||  |  :| | | :|  ||: :|| :|    :|:|| ||:|  | | :|:
Db   412  PGVGVPGVGGAGSGAGSGAGSGAGSGA----GAGSGAGSGAGSGAGSVPGVGVPGVGV   467

Qy   395  PSVG-------GVPGVG-----GVPGVGIGISPEAGAAAAAARAAKYGCTPAAAAAXA----AAX  441
          ||||       |||||      ||| ||||||:|   | :| :|
Db   468  PGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGV-PGVGGAGSGAGSGAGSGAGSGAGS   526

Qy   442  AAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAPGVGPQGVAAA    501
          |  |||| | :|||:|| |||| ||| :||| |  ||| |: :||:| |:|| |
Db   527  GAGAGSVPGVGVGV-  PGVGV-PGVGV-PGVGV-PGVGGAGSGAGSGSG      579

Qy   502  AKSAAKTVAAKAQLRAAAGLGA-------------G:PGLGV-GVGVPGLGV-GAGVPGLGV  547
          |:|| |  | | | | |:|                |:|| || :|  ||| ||:|||| ||
Db   580  AGAGSGAGSGAGSGAGSGAGSGAGSGAGSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGV  639

Qy   548  -GAGVPGF-GAGADEGVRRSLSPELREG-DPSSSGHLPSTPSSPRVPGALAAAAAAAKYGA  604
          ||| :| |  | | |          |   :||||| | :|| |||:|
Db   640  PGVVPGVPGVGG-AGAGSGAGSGAGSGAGSGAGSGAGSVPGVG-VGYPGVGVPGVGYPGV  698

Qy   605  AVPGV-LGGLGALG-----CVGLPG-GVVGAGPAAAAAAAAKAAKAAGFGL----------  648
          |||||  |: ||       ||| || ||| |  ||||||| ||| | |||
Db   699  GVPGVPGVGVSVPGVGVPGVGVPGVGGAGSGAGSGAGSGAGSGAGSGAGSGAGS    758

Qy   649  ------VGAAGLG---GLGVGGLGVPGVG----GLG----GIPTAAAKKAAKYGAAGLG----  690
              ||| | | |    ||| | ||    |||    ::|| | |:|:||:|| | :|:
Db   759  GSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVGGAGSGAGSGSG      818

Qy   691  --GVLGGAGQPPLGGVAARPGFGLSPI-FPG  718
            ||:|| ::||| |  :: |  ::|:| |||
Db   819  GAGAGSGAGSGAGSGAGSGA----GAGSGAGS  849
```

RESULT 5
US-08-212-237-4
; Sequence 4, Application US/08212237
; Patent No. 5605019
; GENERAL INFORMATION:
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Synthetic Proteins As Implantables
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,237
; FILING DATE: 11-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-58847/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-212-237-4

```
Query Match           30.6%;  Score 1159;  DB 1;  Length 832;
Best Local Similarity 44.4%;  Pred. No. 2.4e-70;
Matches 350;  Conservative  49;  Mismatches 243;  Indels 146;  Gaps  55;

Qy     2  GVPG-AIPG-GVPG----GVFYPGAGLGALGGGALGPGGKPLKPVPGGKPLKPVPGGLGAGLGAGLGA  55
          ||||  :|| ||      ||  | :|| | ||  ||:|:|
Db    10  GVPGGVGVPGVPGVGVPGVGVPGVGVPGVGVPGVGV-----AGSSGAGSGAGSGAGS  64

Qy    56  FPAVTFPGALVPGVADAAAAYKAKAGAGLG--GVPGVGGLGVSAGAVVPQ--PGAGV  110
          ||  |||| | || | |||| |   :| :|  | || | :|:|:| |   :|:
Db    65  VPGGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVGGAGSGAGSGAGSSGA  124

Qy   111  KPGKVPGVGLPGV-YPGGVLPGARPGVGVLPGVPT-GAGVKFKAPGVGGAPAG------  162
          |||||||:|  |  ||||| |  ||||| ||| |  :||:|  |  || ||
Db   125  GAGSVPGVGVPGVGVPGVGVPGVGVPGVGV---PGVGGGAGSGAGSGAGSGAGAG  179

Qy   163  -------------IPGVG--PPGGPQPGVP-LGYPIKAPKLPGGYGLPYTGKLPYGYG-  205
                       :||| |  | | |  :|  || |: |:| | ||| | |:|:|:|
Db   180  SGAGSGAGSGAGSVPGVGVPGVGVPGVGVPGVGV---------------GVGVP-  223

Qy   206  PG-GVAGRAGKAG?PTGTGVGPQAAAAAAAAKYGAGAAGV-LPGVG--GAGVPGVPG  261
          || | || | :|| |  | :|:|||||| ||| :| |   || ||  :|||| ||
Db   224  PGVGVPGVGG-AGAGSGAGSGAGSGAGSGAGSVPGVGVPGVGV-PGVGVPGVGV  280

Qy   262  AIPGIGGIAGVGTPAAAAAAAAAKAAKYGAAGLVPGGPGFG--PGVVGVPGAGVPGVG  319
          ||||||:|:|| | |||||||||| ||| :|  |  | | |  || | |:| ||| |
Db   281  GVPGV-GVPGVGVPGVGVPGVGVPGVGVGG----SGAGSGAGSGA-VGVPGVGVPGVG  335

Qy   320  VPGAGIPVVPGAGIPGAAVPGVVPGVSPEAAAKAAAKAAKYGARPGVGVGGIPTIYGVGAGGPP  379
          ||| ||: :|| | || || :| :| | |  :| | | :|  ||:|: |   ::|:| ||:|
Db   336  VPGVG---VPGVGVPGVGVPGVGVPGVGVPGVGGAGSGA---GAGSGAGS----GAGAGSVP  386
```

```
Qy    2 GVPG-AIPG-GVPG-----GVYPGAGLGALGGGALGPGGKPLKPVPGGLAGLAGLGAGLGA  55
          |||| :|| :|| ||||    :|| :| |||| ::|| || | ||| ||| |||| |||| |
Db   10 GVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPG-----AGSGAGSGAGSGAGSGAG  64

Qy   56 FPAVTFPGALVPGGVADAAAYKAKAGAGLG---GVPGVGGLGVSAGAVVPQ---PGAGY 110
          |:| || :||| ||  |  |::|||  |||||    |:| ||||||| |:|    :|:
Db   65 VPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGSGAGSGAGSGA 124

Qy  111 KPGKVPGVGLPGV-TPGGVLPGARPPGVGVLPGVPT-GAGVKPKAPGVGGAFAG--- 162
          | ||||:|| || :|| || | || || |||| |  |||| ||||||||:|:
Db  125 GAGSVPGVGVPGVGVPGVGVPGVGVPGVGV---PGVGGAGSGAGSGAGSGAGSGAG 179

Qy  163 IPGVG-PEGGPQPGVP-LGYPIKAPKLPGGVGLPYTTGKLPYGVG-            205
          : ||| |||| | :|| || ||| ||| || ||: :|:  |:|||
Db  180 SGAGSGAGSGAGSGAGSGAGSGAGSGAGSGA----------GVGVP-------GVGVG 223

Qy  206 PG-GVAGAAGKAGYPGTGVGPQAAAAAAAKAAAKFGAGAAGV-LPGVG---GAGVPGVPG 261
          || |:||:|||  | ||:|| | |  | | || || ||| | |||||    |||||||||
Db  224 PGVGVPGVGG-AGSGAGSGAGSGAGSGAGSGAGSVPGVGVPGVGVPGVGV-- 280

Qy  262 AIPGIGGIAGVGTPAAAAAAAAAAKAAKYGAAAGLVPGGPGFG--PGVGVVPGAGAVGVG 319
          : ||| |||:| | |   | | | | |||| || || |||:|   ||||||||:|||||||
Db  281 GVPGV-GVPGVGVPGVGVPGVGGAGSGAGSGAGSGA---SGAGSGAGSVPG-VGVPGVGVPGVPGVG 335

Qy  320 VPGAGIPVVPGPAGIPGAAVPGVVSPEAAAKAAAKAAAKYGARPGVGVGGIPTYGVGAGGPF 379
          |||:| :|| :|||  | || |   |||| | | | | ||| || :|: ||| :|| | | |
Db  336 VPGVG---VPGVSVGVGVPGVGVPGVG-VAGAGSGA---GAGSGAGSGA---GAGAGSVP 386

Qy  380 GFGVGVGGIPGVPAGVPSVGGVPGVG----GVPGVGISPEAQAAAAAKAAKYGVGIPAAAA 435
          | ||||||:||| |||||:|| |||||    || |||||| |  | | | ||| |||:||||
Db  387 GVGVPGVGVPGV-GVPGV-PGVGVPGVGVPGVGVGV-PGVGV-PGVGV-GAGSGAG 441

Qy  436 AKAAAKAAQFGLVPGGVVGGLAPGVGVAPGVGLADPGVGVAPGVGLAPGVGVAPGVGIGP 495
          :| || |  |||| || ||||||||||||||| | |||||||||||:| ||||||||:|:
Db  442 SGAGA------GSVPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVGG 489

Qy  496 GGVAAAKSAAKYAAKAQLRAAAGLIAGIPGLGV-GVGVPGLGV-GAGVPGLGV-GAGVP 552
          :|||| | || | | | | | | |||| |||:|  :||:|| | ||| |:| | |:| ||
Db  490 AGAGSGAGSGAGSGAGSGAGSVPGVGVPGVGVPGVGVPGVGVPGVGVGVGVGVP 549

Qy  553 GF-GAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGV-L 610
          || ||| ||:| ::: | | | |||:||| | |   || || | | | |:|:| |||||||
Db  550 GVGGAGSGAGSGAGSGAGSGAGS-----------VPG-VGVPGVGVPGVGVPGVGV 597

Qy  611 GGLGALG---GVGIPG-GVVGGLVGAAGPGLVGAAGLGLGVGGLGVP 665
          :| | |:|    ||:|| |||||||||||||| || || | :|:|||
Db  598 PGVGVPGVGVPGVGVPGVGVGGAGSGAGSGAGSGAGSGAGSGAGSVPGVGVPGVGVP 657

Qy  666 GVG------GLG----GIPPAAAAKAAKYGAAGLGVLG---GAGQPPLGGVAARPGFG 711
          |||      |||     ||| | | | | | |||  |||   |:||| :|||||:| |
Db  658 GVPGVGVPGVGVPGVGVPGVGVPGVGGAGSGAGSGAGSGAGSVPGVGVGV---PGVG 714

Qy  712 LSPI-FPG 718
          : :| ||
Db  715 VPGVGVPG 722
```

COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/212,237
  FILING DATE: 11-MAR-1994
  CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
  NAME: Rowland, Bertram I
  REGISTRATION NUMBER: 20,015
  REFERENCE/DOCKET NUMBER: A-58847/BIR
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 415-781-1989
  TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 3:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 936 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: protein

Query Match   29.9%; Score 1130; DB 1; Length 936;
Best Local Similarity 38.5%; Pred. No. 2.3e-68;
Matches 377; Conservative 46; Mismatches 248; Indels 308; Gaps 62;

```
Qy    2 GVPG-AIPG-GVPG----GVFYPGAGLGALGGGALGPG------------------------ 33
          ||| :|| |||| |      :||
Db   10 GVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGSGAGSVPGVGVPGVGVPGVGVP 69

Qy   34 -----------GKPLKPVPG-GLAGAGLGAGLGAPAVTFPGALVPG------------ 68
                      || :||||| ||||||| || ||
Db   70 GVGVPGVGVPGVGVPGVGGAGSGAGSVPGVGVPGVGVPGVGVPGVGVG 129

Qy   69 ------GVADAAAAYKAAKAGAGLGVPGVG-------GLGVSA----GAVVP---QPG 107
                :|| | ||||||||                :|| |      ||||
Db  130 VPGVGVPGVGVPGVGGAGSGAGSVPGVGVPGVGVPGVGVPGVGVPGVGVPG 189

Qy  108 AGVK----------------DGKVPGVGLPGV-YPDGVLPGARPPGVGVLPGVTP-GAGVTKPKA 153
          |:||                || :|| || : :||:|
Db  190 VGVPGVGGAGSGAGSVPGVGVPGVGVPGVGVPGVGVGV----PGVGVPGVGV---- 244

Qy  154 PGVGGAPAG-------IPCVGPPGPGPGPQPGVPL-GYPIKAPKLPG----GYGLPVTTGKLP 201
          ||||| || |      |:|| |:::||||||||||  :|
Db  245 PGVGGAGSGAGSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGA- 302

Qy  202 YGTGPGGGVGAAGKAGYP----------------TGTGVGPQAAAAAAAKAAAKTGA 242
          ||| || |||:|||||||                |:||| |
Db  303 -GAGSGAGSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVGGAGSGA 360

Qy  243 GAAGVLPGVG--GAGVP--GVPG-AIPGIG---GIAGVGTPAAAAAAAAKAAAKTGAA 293
          |:| || |:|  |||:|  |||  ||||      || |
Db  361 G-AGSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGAGS------GAG 413

Qy  294 AGLVP--GGPGPG-PGV---VGVPGAGVPGVGVPGVGAGIP------------------ 326
          || || ||:|:|  ||:   |||||||||||||||||||| :|
Db  414 AGSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVGGAGAGSGAGSVPGVG 473

Qy  327 -----------VVPGAGIPGAAVPGTVSPEAAAAAAKAAKYGARPGVVG---GIPT 370
                     :|| |:|| || :|| ||:|||||||||||:||| ::|||
Db  474 VPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGSGAGSVPGVGVPGVGVPG 533

Qy  371 YGVGAGGPGPGV-GVG----GIPGV-------------------AGVPSVG----GVPGV- 404
          || || ||:|:| |:|    |:||: |                 ||:||| |    :||||
Db  534 VGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGSGAGSVPGVGVPGVGVPGVGVGV 593

Qy  405 ---GVFPGVGISPEAQAAAAAKAAKYGVGTPAAAAAKAAAKAAAQPGLVPGVGVPGVGVAP 461
             :|| ||||            :|||   |:|||||||||:|||||| ||::||||||||| |
Db  594 PGVVPGVGVP--------GVGVPSVGA-GAGSGAGSVPGVGVPGVGVPGVGVGV-P 636
```

```
Qy  462 GVGVAPGVGLAPGVGVAPGVGVAPGVGVAPGIGPGGGVAAAAKSRAKYAAKAQLRAAAGLG 521
          |||| |||| |||| ||||| |||||| ||: ||||                  ::| :
Db  637 GVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVGGAGAGSGA----------GAGSVPG 679

Qy  522 AGIPGLGV-GVGVPGLGV-GAGVPGLGV-GAGVPGF-GAGADEGVRRSLSPELREGDPSS 577
          |:||:|| ||||||||  :|||||||  :||||| |:||:||||              :|:
Db  680 VGVPGVGVPGVGVPGVGVPGVGVPGVGVGGAGSGAGSVPGV--GVPGV 737

Qy  578 SQHLPSTP-------------------SSPRVPGALA--AAKAAKYGAAVPGV-LGGLGAL 616
          ||::|::|                    :||||| |:  |||||||||:|||| :|:||:|:
Db  738 GVPGVGVPGVGVPGVGVPGVGVPGVGGAGSGAGSVPGVGVGVPGVGVGVP 797

Qy  617 G----GVGIPG-GVVGAG-PAAAAAAKAAKAAAQFGLVGAAGLG------------- 655
          |    |||:||  :|| |  |||||||||||||::|:||||||:
Db  798 GVGVPGVGVPGVGVPGVGGAGSGAGSVPGVGVPGVGVPGVGVPGVGVG 857

Qy  656 --GLGVGGLGVPGVGGLGGIPPAAAAKAAKYGAAGLG----GVLG-GAGQPLGGVAARP 708
            |:||||||||||||:||| ||||||||||||||||    ||:|  |||:||||||| |
Db  858 VPGVGVPGVGVPGVGGAGSGAGSVPGVGVPGVGVPGVGVGVPGVGVGV-P 916

Qy  709 GPGLSPIPPGGACLGKACG 727
          |:|::|:||||| ||:| |
Db  917 GVGVPGV--GGAGAGSGAG 933
```

RESULT 8
PCT-US95-02772-3
; Sequence 3, Application PC/IUS9502772
; GENERAL INFORMATION:
  APPLICANT: Protein Polymer Technologies, Inc.
  TITLE OF INVENTION: Synthetic Proteins As Implantables
  NUMBER OF SEQUENCES: 9
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
    STREET: Four Embarcadero Center, Suite 3400
    CITY: San Francisco
    STATE: CA
    COUNTRY: U.S.A.
    ZIP: 94111-4187
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: PCT/US95/02772
    FILING DATE:
  ATTORNEY/AGENT INFORMATION:
    NAME: Rowland, Bertram I
    REGISTRATION NUMBER: 20,015
    REFERENCE/DOCKET NUMBER: FP-58847-1-PC/BIR
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 415-781-1989
    TELEFAX: 415-398-3249
  INFORMATION FOR SEQ ID NO: 3:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 936 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    MOLECULE TYPE: protein
PCT-US95-02772-3

Query Match   29.9%; Score 1130; DB 5; Length 936;
Best Local Similarity 38.5%; Pred. No. 2.3e-68;
Matches 377; Conservative 46; Mismatches 248; Indels 308; Gaps 62;

```
Qy    2 GVFPG-AIPG-GVPG----GVFYPGAGLGALGGGALGPG------ 33
          |||| :|| ||||         :||
Db   10 GVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGSGAGSVPGVGVPGVGVPGVGVP 69
```

```
TITLE OF INVENTION:  OF ENZYMATIC CROSS-LINKING
NUMBER OF SEQUENCES: 105
CORRESPONDENCE ADDRESS:
  ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
  STREET: 4 Embarcadero Center, Suite 3400
  CITY: San Francisco
  STATE: California
  COUNTRY: USA
  ZIP: 94111-4187
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/397,633A
  FILING DATE:
  CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
  NAME: Rowland, Bertram I
  REGISTRATION NUMBER: 20,015
  REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (415) 781-1989
  TELEFAX: (415) 398-3249
  TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 54:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 877 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-397-633A-54
```

```
Query Match             29.8%;  Score 1126.5;  DB 1;  Length 877;
Best Local Similarity   44.0%;  Pred. No. 3.7e-68;
Matches  347;  Conservative  59;  Mismatches  230;  Indels  153;  Gaps  57;

Qy    2 GVPG-AIPG-GVPG--------GVFYPGAGLGALGGALGPGGKPLKPVPGGLAGAGLGAGLG-  54
            |||   ||  ||||        ||  |   |:| | :|||  |:||||  |:|:|
Db   84 GVPGVGVPGVGVPGVGVPGVGVPGVGVPGAGAGS-GAGA-GSGA------GAGSGAGAGSGVGV 134

Qy   55 ------AFPAVTFPGALVPG----GVADAAAAYKAAKAGAGLGGVPGVG-GLGVSAGA     101
              | |  |  |||| |||     ||        |:| | ||||||:||| :|:|  |
Db  135 PCGVGVPGVGVPGVGVPGVGVPGVGVPGVGVGLPGV-YPGGVLPGARPPGVGVTPGAGV----- 194

Qy  102 VVPQFGAGVFPGKPGVGVGVLPGV-YPGGVLPGARPPGVGVLPGVPTGAGV-KPKAPGVG--   157
             |    |  ||| || :| || |  ||||| |||||||| |||||:||| |||||||
Db  195 GVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGV-PGAGAGSGAGAGASGG   253

Qy  158 ---GAFAGIPGVPGQPGVPGSVPLGYPIKAPKLPGGYGLPYTTGKLPYGYG--PG--GVAGA 212
               |     |  |||  | :| :| ||  | | | | ::||||  | | |||     ||  ||
Db  254 AGAGSSGVSGVPGVGVPGVGVGV--GVP--------------GVGVP-------GVGVPGVG-  294

Qy  213 AGKAGYFPCGCVGFQAAAAAAAAKAAKFPGAGAAGY-LPGVG--GAGYPGVPGCAIPGIG--  267
           || |  | |    | |||| ||| |   | | ||     || || |||  |  |||  |
Db  295 VGTPGAGAGSGAGSGAGSGAGASGSGVPGVGVPGVGVPGVGVPGVGVPGV--GVPGVGVP----- 352

Qy  268 --GIAGVGTPAAAAAAKAAAAKAAKYGAAGLVPGGPGPGFGVVGPGVGVPGRAGI        325
              | :||| | || ||| || |||| ||   |||||| || :|::||||| |||::
Db  353 GVGVPGVGVPCAGAGSGAGAGAGS--GAGAG--SGAGASG-VGTPGVGVPGVGVPGVGVP----- 405

Qy  326 P--VVPGAGIPGAAVPGVVSPEAAKAKAAAKAAAXYGARPGVGVGLPPTVGVGAG---GFPG   380
              |    |  | ||| | ::| || ||||| |||| |||||  |:||  :||  ||     |||
Db  406 PCGVGVGVGVPGVGVPGVGVPGVGVPGAGAGSGA-----GAGSGAGAGS---GAGAGSGVGVPG  456

Qy  381 FGVGVGGIPGVAGPSVGGVPGVG----GVPGVGISPEAQAAAAKAAKYGVGIPAAAAA      436
             | | | |    || ||||  ||    ||||||  | :| || |||  |  |||||||
Db  457 VGVPGVGVPGV-GVPGV-GVPGVGVPGVGVPGVGVPGVGV--PGAGAGSGA-----GAGSGAGS  508

Qy  437 KAAAKAAQFGLVPGVGVAPGVGVAPGVGVGVAPGVGVGVAPGVGLAPGVGVAGVGVAPGIGPG  496
```

```
Qy   34 --------GKPLKPVPG-GLAGAGLGAGLGAPPAVTFPGALVPG---------         68
                  |  |  :|||   | ||||| |: :  : | | : | | |:
Db   70 GVGVPGVGVPGVGVPGVGGAGAGSGAGAGSGAGAGSGVPGVGVPGVGVPGVGVPGVG        129

Qy   69 --------GVADAAAAYKAAKAGAGLGGVG---------GLGVSA---OPG          107
                  |  | :::|  | | ||||| ||         |:|:||    ||
Db  130 VPGVGVPGVGVGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVPG                    189

Qy  108 AGYX--------PGKVPGVGLPGV-YPGGVLPGARFPGVGVLPGVPPT-GAGVKPKPKA     153
             ||         :||| |  ||| |||| |||  |||||| |||:|| |  |: |  |
Db  190 VGPVGPGVGGAGSGAGAGSGAGAGSGAGSGAGSGVPGVGVPGVGVGV-PGVGVPGVGV----  244

Qy  154 PGVGGAFAG------IPGVGPFGGPGQPGVPL-GYPIKAPKLPG----GYGLPYTTGKLP    201
             | || | ||     |||| |  || ||  |   ||: | | ||     |:|| |: |||
Db  245 PGVGGAGAGSGAGAGSGVPGVGVPGVGVPGVGVPGVGVPGVGVGVGV-PGVGVPGVGGA----  302

Qy  202 TGTGPGGVAGAGAKGYP---------               242
             |:| || | |  |: |
Db  303 -GAGSGAGAGSGVPGVGVPGVGVPGVGVPGVGVPGVPQAAAAAKAAAKFGA              360

Qy  243 GAAGVLPGVG---GAGVP---GVPG-AIPGIG-----GIAGVGTPAAAAAAKAAKYGAA    293
             | | |||||   | | |   | ||  |||:|     |:||||||  |||| ||||:  |
Db  361 G-AGSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVGV-----GAG        413

Qy  294 AGLVP---GGPGFG-PGV----VGVPGAGVPGVGVPGAGIP---------          326
             ||:|    | |:| | |    |||| ||| | ||| ||  |
Db  414 AGSVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGAGSGAGAGSVPGVG         473

Qy  327 ----VVPGAGIPGAAVPGVVSPEAAKAKAAKYKAAXYGARPGVGVG--GIPT        370
                 |||| || ||  ||:| :| ||| | ||  | | |||||||||  |:|
Db  474 YPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGAGSGAGAGSVPGVGVPGVPG      533

Qy  371 YGVGAGGPGPGRGV-GVG-------AGYPSVG----GVPGGVG-          404
             :|||  | | | :| ||:        |:| |     |:||  |
Db  534 VGVPGVGVPGVGVPGVGGAGAGSGAGAGSGAGSGAGSVPGVGVPGVGVPGVG            593

Qy  405 ---GVPGCVGISPEAQAAAAKAAAKAAKYGVGTPAAAAAKAAAAQFGLVPGVGVVAPGVAP  461
                |:| |    | :| || ||| ||| :| |  |||| || | ::| |:|: |  |
Db  594 PGVGVPGVGGA-GAGSGAGAGSGAGAGSGVPGVGVGVPGV-PG--GVAGA---------   636

Qy  462 GVGVAPGVGLAPGVGVAPGVGVAPGIGPGVAAAAKSAAKVAAKAQLRAAAGLG         521
             |:||  |  | |    | |::| |  || |  | ::||| ||  ||  |:  |
Db  637 GVGV-PGVGV-PGVGV-PGVGVPGVGVPGAGAGSGA-----------GAGSVPG        679

Qy  522 AGIPGLGV-GVGVPGLGV-GAGVPGF-GAGADEVRRSLSPELREGDPSS          577
             ||:| :| | || | :| | | ||  |:| | | : | :|  : |
Db  680 VGVPGVGVPGVGVPGVGVPGVGGAGAGSGAGAGSGAGSVPGV--GVPGV             737

Qy  578 SQHLPSTP-----------------SSPPVPGALA----AAKAKYGAAVPGV-LGGLGAL   616
              |  |:||                  : | |||||    | || | || ||| |  | |
Db  738 GVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVPGVGVPGVP   797

Qy  617 G----GVGLPG-GVVGAG-PAAAAAAKAAAKAAQFGLVGAAGLG------          655
                  |:|:||  | || |  || || ||| |:||:| ||| | |:|
Db  798 GVGVPGVGVPGVGVPGVGGAGAGSGAGAGSGAGSVPGVGVPGVGV         857

Qy  656 --GLGVGGLGVPGGLGIPPAAAAAKAAKYGAAGLG----GVLG-GAGQFPLGGVVARP    708
               |:|:|  | ||| |  |||| || |||| | |    |: | | | |  |:||: |
Db  858 VGVPGAGAGSGAGSGAGSGAGSGVPGVGVPGVGVPGV------GVPGVGVPGVGVPGVGVP-   916

Qy  709 GFGLSPIFPGGAGLGKACG 727
             | | ::| | | ||  |
Db  917 GVGVPGV--GGAGAGSGAG 933
```

```
Db    509 GAGAGS---------GVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVGVP 554
              |:|       |||  ||||  ||||  ||||  ||||  |||| |||:
Qy    497 GVAAAAKSAAKVAAKAAQLRAAAGLSAGIPGLGV-----GVGVPGLGV-GAGVPGLGV-G 548
              |:|  :|  :| :  |: ||| |  ||:|:|:||  ! |||:||:|| |  ! |:|
Db    555 GAGAGSGAGAGSGAGAGSGAGAGSGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPG 614

Qy    549 AGVPGFGAGADEGVRRSLSPELREDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPG 608
              |||  |:|:         :       |  :  ::|       |  |:|
Db    615 VGVPGAGAGSGAGA---------GSGAGSGAGAGSGVGVPG-VGVPGVGVPGVGVPG 663

Qy    609 V-LGGLGALG----GVGIPGGVVGAGPAAAAAAKAAAKAAQPGL---VGAAGLG--GL 657
              : |::|  |:|  ||||| |:|  ||:  :  :|   || ! ||:|  : |: :
Db    664 VGVPGSVGVPGVGVPGVGVPGVGPGAGAGSGAGAGSGAGAGSGVGVPGVGVPGVGVPGV 723

Qy    658 GVGGLGVPGVG--GLG----GIPPAAAAAKAAKYGA-AGLGGVLGGAGQFPLGGVAARGF 710
              || |:||||||  |:|       :|:      |:| ||:||:||:| ! |: |
Db    724 GVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGAGAGSGAGSGVGVPGVGVPGV-PGV 782

Qy    711 GLSPI-PPG 718
              |:| |: |
Db    783 GVPGVGVPG 791
```

RESULT 10
US-08-397-633A-68
; Sequence 68, Application US/083976233A
; Patent No. 5773577
GENERAL INFORMATION:
  APPLICANT: Cappello, Joseph
  TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE
  TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING
  NUMBER OF SEQUENCES: 105
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
    STREET: 4 Embarcadero Center, Suite 3400
    CITY: San Francisco
    STATE: California
    COUNTRY: USA
    ZIP: 94111-4187
  COMPUTER READABLE FORM:
    MEDIUM TYPE: floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/397,633A
    FILING DATE:
    CLASSIFICATION: 530
  ATTORNEY/AGENT INFORMATION:
    NAME: Rowland, Bertram I
    REGISTRATION NUMBER: 20,015
    REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (415) 781-1989
    TELEFAX: (415) 398-3249
    TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 68:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 884 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: protein
US-08-397-633A-68

```
Query Match    29.7%;    Score 1123;    DB 1;    Length 884;
Best Local Similarity  41.8%;  Pred. No. 6.4e-68;
Matches 352; Conservative  53; Mismatches 249; Indels 188; Gaps  56;

Qy     2 GVPG-AIPG-GVPG----GVFVPGGLGALGGGALGPG--------GKPLKPVPG-G 43
              |||| :||:|||       || |||| |:  |: |         | | |||| |
```

```
Db    61 GVPGKGVPGVGVPGVGVPGVGVPGAGAGSGAGAGSGAGAGSGAGAGSGVGVPGVGVPGVG 120
              |||  ||||  ||||  |||| |||:| |||| ! |:|   ----GVPGVG 93
Qy    44 LAGAGLGA---GLGAFPAVTFPGALVPG-GVADAAAAYKAAKAAGAGLG----GVPGVG 93
              : |||   |||    | ||| |:|||:|  || |:|:       !  |||: !
Db   121 VPGVGVPGKGVPGVGV-PGVGVPGVGVPGAGSGAGSGCAGSGAGSGAGSGVGVPGVGVPGVG 179

Qy    94 --GLGYSAGAVVPQPGAGVKPGK-VPGVGLPGVYPGGVYLPGARFPGVGVLPGVPTGAGVK 150
              :|||:||  |   |:||   ||: ||||  |:||||:||:|||   !
Db   180 VPGVGV-------PGVGV-PGKGVPGVGVPGY---GV-PGVGVPGAGSGAGAGSGAG 226

Qy   151 PKAPGVSGAFAGIPGVGPFGG2QPGVPLGYPIKAPKLPGVGLPYTTGKLPYGYGPGGVA 210
              |:| ! |||:|  : ||| |: !  :|| ! :| :||| | :| ||:| ! |
Db   227 AGSGAGAGSGVGVPGVGVPGVGVPGV--GVP--------GKGVP-------GVGVPGV- 267

Qy   211 GAAGKAGYPTGTGVGPQAAAAAAAKAGAAGVLPGVG---GAGVP--GVPG-AIPG 265
              || ! |:|: | || ! | !  || ||  | |:| ! |:|| !  |||| |:|
Db   268 GVPG--VGVP-GAGAGSGAGSSGAGAGSGVGVPGVGVPGVGVPGVGVPGKGVPG 325

Qy   266 IG----GIAGVGTPAAAAAAAKAAKYGAAAGLVPGGPGFGPGVPGAGVPGVGVP 321
              |:  |: |||||||:|:|:|:|:|||||| ||| ! ||||||:|||||||||||
Db   326 VGVPGVGVPGVGVPGAGAGSGAGAGS--GAGAG--SGAGAGSG-VGVPGVGVPGVGVP 378

Qy   322 GAGIP--VVPGAGIPGAAVPGVVSPEAAAKAAKA----AKYGARPGVVG-----GIPT 370
              ! |: ! |||| |:  ! ::|||  |: || |: ! !  |:|| |:|:|:     ! |
Db   379 GVGVPGKGVPGVGVPGVGVPGVGVPGAGAGSGAGAGSGAGSGAGSGVGVPGVGVPG 438

Qy   371 YGVGAGGFPGFGV-GVG----GIPGVAGVPSVG---------GVPG 402
              ||  !  !  || !  :||    || ||| |:|          |||
Db   439 VGVPGVGVPGKGVPGVGVPGVGVPGV-GVPGAGAGSGAGAGSGAGSGVGVPGVGVPG 497

Qy   403 VG---GVPGVGISPEAQAAAAAAKAAKY-GVGTPAAAAAKAAKAAQPGLVPGVGVAPGV 457
              ||     |||| ||:|   :     |:| |||:|||:|:|:|:|||||||:|||||:|
Db   498 VGVPGVGVPGVGVPGKGVPGVGVPGVGVPGVPGAGAGSGAGSGAGAGSGAGSGVGVPGV 557

Qy   458 GVAPGVAPGVGVGLAPGVGV-----APGVGVAPGVGVARGIGPGGVAAAAKSAAKVAAKAQ 513
              || ||| ||||:| ||| !  !     ||| ||:||||| !  : :   |||||||:|
Db   558 GV-PGVGV-PGVGV-PGVGVPGKGVPGVGV--PGVGV-PGVGVPGAGAGSGAGSGAGSGAG 612

Qy   514 LRAAAGLGAGIPGLGV-GVGVPGLGV-----GAGVPGLGV-GAGVPGFGAGADEGVRRS 565
              ! | |:|:|||| ! |:||||:||:|     ! |:|:|:| ! |:|:|:| |:|:| !
Db   613 SGAGAGSGVGVPGVGV-PGVGVPGVGVPGKGVPGVGVPGVGVPGAGAGSGAGA--- 669

Qy   566 LSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGV-LGGLGALG----GVG 620
              !  |:|:| !  : |  : ! : : |||  |:|:|:|  ||| |:|| |:| !   ||:
Db   670 ------GSGAGAGSGAGSGVGVPG------VGVPGVGVPGKGVPGVGVPGVGVPG 716

Qy   621 IPG-GVVGAGPAAAAAAKAAKAAQPGLVGAAGLGGLGVGGLGVPGVG--------- 668
              ||| ! || ||| !  :|:|:|:|||||:|| |:|:|:||||:|||
Db   717 VPGVGVPGAGAGSGAGSGAGSGVGVPGVGVPGVGVPGKGVPGVGVGV 776

Qy   669 -GLG----GIPPAAAAAAKYGAAGLGGVLGGAGQFPLGGVAARPGFGLSPI-FPGGACL 722
              ! |:|    ! :|   |:|:|  ||:|| |:| ! : ||:|: |: : |:|! |||| |
Db   777 PGVGVPGVGVPGAGAGSGAGSGAGAGSGAGAG---SGVGV-PGVGVPGVGVPGVGVP 831

Qy   723 GK 724
              ||
Db   832 GK 833
```

RESULT 11
US-08-435-641-15
; Sequence 15, Application US/08435641
; Patent No. 5817303
GENERAL INFORMATION:
  APPLICANT: Stedronsky, Erwin R.
  TITLE OF INVENTION: Tissue Adhesive Using Synthetic
  TITLE OF INVENTION: Crosslinking
  NUMBER OF SEQUENCES: 15
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
    STREET: Four Embarcadero Center, Suite 3400
    CITY: San Francisco
    STATE: CA

```
          COUNTRY: US
          ZIP: 94111
      COMPUTER READABLE FORM:
          MEDIUM TYPE: Floppy disk
          COMPUTER: IBM PC compatible
          OPERATING SYSTEM: PC-DOS/MS-DOS
          SOFTWARE: PatentIn Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
          APPLICATION NUMBER: US/08/435,641
          FILING DATE: 05-MAY-1995
          CLASSIFICATION: 514
      ATTORNEY/AGENT INFORMATION:
          NAME: Rowland, Bertram I.
          REGISTRATION NUMBER: 20015
          REFERENCE/DOCKET NUMBER: A-61127
      TELECOMMUNICATION INFORMATION:
          TELEPHONE: 415-781-1989
          TELEFAX: 415-398-3249

  INFORMATION FOR SEQ ID NO: 15:
      SEQUENCE CHARACTERISTICS:
          LENGTH: 884 amino acids
          TYPE: amino acid
          STRANDEDNESS: single
          TOPOLOGY: linear
          MOLECULE TYPE: peptide
  US-08-435-641-15

  Query Match           29.7%;  Score 1123;  DB 2;  Length 884;
  Best Local Similarity 41.8%;  Pred. No. 6.4e-68;
  Matches  352;  Conservative  53;  Mismatches  249;  Indels  188;  Gaps  56;

Qy    2  GVPG-AIPG-GVPG------GVPYPGAGLGALGGGALGPG--------GKPLXPVPG-G   43
            ||| .:|| ||||      | || ||| || ::   | |        :|  | ||| |
Db   61  GVPGKGVPGPGVPGVPGVPGPGAGAGSGAGSGAGSGAGSGVPGVPGVPGVPGVPGVG  120

Qy   44  LAGAGLGA----GLGAFPAVFTPGALVPG-GVADAAAAYKAAKAGAGLG------GVPGVG   93
           :.|.:|      || | :  | | |  |  :  |  | | :|:.  | |       | ||
Db  121  VPGVGVPGKGVPGVPGGV-...                                       179

Qy   94  --GLGVSAGAVVPGPGRAGVTKPGK-VPGVGLPGVTPGGVLPGGARPGVGVLPGVPTCGAGVK  150
           :|| :  | |  | |  | : | ||  | :: |.| || | |  |:|  |
Db  180  VPGVGV-----PGVGV-PGKGVPGVGVPGV--GV-PGVGVPGAGAGSGAGSGAGSGAG       226

Qy  151  PKAPGVGGAFPAGIPGVPGGPGPGVPLGVPIKAPKLPGGIGLPYTTGKLPYGVPGPGGVA   210
             | ||  | | || |||| |  |  | | |:|||||| : ::  ||| :| |
Db  227  AGSGAGAGSGVGVPGVGVPGVGVPGV--GVP------GKGVP------GVGVPGV-      267

Qy  211  GAAGKACYPTGTGVGPQRAAAAAAAKAAAKFGAGAAGVLPGVG---GAGVP-GVPG-ALPG   265
             | :|  |    | || |  | | |||| : |    :|: ||    ||| | | |  |||
Db  268  GVPG-VGVP-GAGAGSGAGSGAGSGVGVPGVGVPGVPGKGVPG                   325

Qy  266  IG----GIAGVGTPAAAAAAAAAKAAAKYGAAGLVPGGPGPGPGVGVGVPGGAGVGVP     321
          :|    |:||||  |  | | ||| : |:|  | : || | | |  |    |  | ||
Db  326  VGVPGVPGVGVPGAGSGAGSGAGAGS---GAGAG---SGAGASG-VGVPGVPGVPGVGVP   378

Qy  322  GAGIP--VVPGAGIPGAAVPGVVSPEAAAKAAAKA----AKYGARPGVGVG-----GIPT   370
           ||:|   | || |||  |||   |  | | ||| |     | | | | |      ||:|
Db  379  GVGVPGKGVPGVGVPGVPGVGVPGAGSGAGSGAGAGSGVPGVGVPG                438

Qy  371  YGVGAGGFPGFGV-GVG----GIPGVAGVPSVG------------------GVPG       402
          :|||| | || ||  | |    ||||  |||  |                  ||||
Db  439  VGVPGVGVPCKGVPGVPGVPGV-GVPGAGSGAGSGAGAGSGVGVPGVPG            497

Qy  403  YG----GVPCGVGISPEAQAAAAAAKAKY-GVGTPAAAAAKALAKAAAQFGLVPGVTAPGV   457
          :|    |||  ||     | | | |||:|   :.|  | | ||| :|  ||| || | ||
Db  498  VGVPGVGVPGVPGKGVPGVPGVPGVGVPGAGSGAGSGAGAGSGAGAGSGV           557

Qy  458  GVAPGVGAPGVAPGVGVGV----APGVGVAPGVGVAPGIGPGVAAAAKSAAKVAAKAQ     513
           || ||| |||  || | |    ||| || |||||||| |.| | | :|:||:||||:|
Db  558  GV-PGVGV-PGVGV-PGVGVPGKGVPGVGV-PGVGVPGAGSGAGSGAGSGAGAG        612

Qy  514  LRAAAGLGAGIPGLGV-CVGVPGLGV------GAGVPGLGV-GAGVPGFGAGADEGVRRS   565
```

```
Db  613  SGAGAGSGVGPGVGPGVGPGVGPGKGVPGVPGVPGVGPGVGPGVGPGVGAGSGAGA---   669
          || |||| | ||: |||: |: :|  |  | ||||: || |.| |: |  |  |||
Qy  566  LSPELRESDPSSQHLPSTPSSPRVPGALAAAKAAKYGAVPGV-LGGLGALG---GVG    620
Db  670  ------GSGAGAGSGAGSGVGVPG------VGVPGVPGVPGVPGKGVPGVGVPGVGVG   716
Qy  621  IPG-GVVGAGPAAAAAAKAAAKAAQFGLVGAAGLGLGVGGLGVPGVG---------    668
          :||:||||| |  | | ||| | | ||| | | |:|  ::| | |
Db  717  VPGVGVPGAGAGSGAGSGAGSGAGSGVGPGVGPGVPGVPGKGVPGVGV          775
Qy  669  -GLG----GIPPAAAKAAKYGAAGLGGVLGGAGQFPLGVAARPGFGLSPI-FPGGACL   722
           ||     ::   | | |||  | |: ::  |||   || || ||| : :  : .  :|
Db  777  PGVGVPGVPGAGSGAGSGAGSGAGSGAGAG----SGVGV-PGVGVPGVGVPGVGVP    831
Qy  723  GK                                                         724
          ||
Db  832  GK                                                         833
```

```
;       TELEX:   910 277299
; INFORMATION FOR SEQ ID NO:  96:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH:  884 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;       MOLECULE TYPE:  protein
US-08-707-237A-96
```

```
Query Match          29.7%;  Score 1123;  DB 2;  Length 884;
Best Local Similarity  41.8%;  Pred. No. 6.4e-68;
Matches  352;  Conservative  53;  Mismatches  249;  Indels  188;  Gaps  56;

Qy      2  GVPG-AIPG-GVPG----GVFYPGAGLGALGGGALGPG----------GKPLKVPG-G   43
              :||| |||| :||  | || : ||  |  :||| |||          :|:| :|:| |
Db     61  GVPGKGVPGVGVPGVGVGVPGVGVGVPGVGAGSSGAGAGSSGAGAGSGVGVPGVGVPGVG  120

Qy     44  LAGAGLGA-----GLGAFPAVTFPGALVPG-GVADAAAYKAAKAGAGLG------GVPGVG   93
              ||| |||:    |  ::  | || | |  ||  ||   :||: |   | |       || ||
Db    121  VPGVGVPGKGVPGVGV-PGVGVPGVGVGLPGVGVLPGGVLPGARPPGVGVLPGVPTGAGVK  179

Qy     94  --GLGVSAGAVVPQPGAGVKPGK-VPGVGLPGVTPGGVLPGARPPGVGVLPGVPTGAGVK  150
             | || : | | ||  |  | ||   |||| :  |  |  ||| || |||| | :
Db    180  VPGVGV-------PGVGV-PGKGVPGVGVPGV--GV-PGVGVPGAGAGSSGAGAGSSGAG  226

Qy    151  PKAPGVGGAPAGIPGVGPPGGPGPQPGVPLGYPIKAPKLPGGLPYTTGKLPYGYGPGGVA  210
              | | | |  : |   |  || |  |  | :  |||:  |: | | ||  |  | | | |
Db    227  AGSSGAGAGSSGVPGVGVPGVGVPGV--GVP-----GKGVP--------GVGVPGV-   267

Qy    211  GAAGKAGYPTGTGVGPGQPAAAAAKAAAKPGAGAAGYLPGVG--GAGVP--GVPG-AIPG   265
              ||| |   | || | |   ::: : |||:  ||  |  ||  | | |  || :| | |
Db    268  GVPG--VGVP-GAGAGSSGAGAGSGAGSGVGVPGVGVPGVGVGVPGVPGKGVPG   325

Qy    266  IG-----GIAGVGTPAAAAAAAAAAAKAAKYGAAAGLVPGGPGPGPGVGVGAGVPGVGVP  321
              :|     | :| |   || | | |  |||  :  |  |: || || | || |  |||  |
Db    326  VGVPGVGVPGVGVPGAGAGSSGAGAGS--GAGAG-SGAGAG-SGAGAGS-GVGVPGVGVP  378

Qy    322  GAGIP--VVPGAGIPGAAVPGVVSPEAAAAKAAKA-----AKYGARPGVVG-----GIP   370
              | |:|   | || || |  |   ::   ||: |  |     ||  | |||      |:
Db    379  GVGVPGKGVPGVGVPGVGVPGVGVPGAGAGSSGAGAGSSGVGVPGVGVPG   438

Qy    371  YGVGAGGPGPGGV-GVG------GIPGVAGVPSVG----------------GVPG   402
              | | | |  |  | |||      ||  | | | |                  ||||
Db    439  VGVPGVGVPGKGVPGVGVPGVGVPGVGVPGAGAGSSGAGAGSSGAGAGSGVGVPG   497

Qy    403  VG-----GVPGVGISPEACAAAAAKAAAKY-GVGTPAAAAAKAAAKAAQFGLVPGVGVAPGV  457
              ||     |||  | :  |  : |||: ||   |||  | |  :|  ||| | | | |
Db    498  VGVPGVGVPGKGVPGVGVPGVGVPGVGVPGAGAGSSGAGAGSSGAGAGSG   557

Qy    458  GVAPGVGVAPGVGVLAPGVGV----APGVGVAPGVGVGPGVGVAAAKSAAKVAAKAQ  513
             |  | ||  || ||||  |: |    ||| | | || || | |  | | :|| | :|:
Db    558  GV-PGVGV-PGVGV-PGKGVPGVGVPGVGVGVPGVGVPGVGVPGAGAGSGAG   612

Qy    514  LRAAAGLGAGIPGLGV-GVPGPGLGV------GAGVPGLGV-GAGVPGFGAGADEGVRRS  565
              | | | | | || ||  || |  ||       | | || ||  | | || | | |  |
Db    613  SGAGAGSGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGAGAGSSGAGA--   669

Qy    566  LSPELREGDFSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGV-LGGLGALG----GVG  620
              |      | :      | | | : ||  |  :||:|   | ||  | | |||     || ||
Db    670  G------GSSGAGAGSSGVGVPGVGVPG------VGVPGVGVPGKGVPGVGVPGVG   716

Qy    621  IPG-GVVGAGPAAPAGIPGVGPGGPQPGVPLGYPIKAPKLPGGLPYTTGKLPYGYGPGGVA   668
              :|| || |  |  |   | |  | | |  |  | :|||:  |: |  |||  || | | |
Db    717  VPGVGVPGAGAGSSGAGAGSSGVGVPGVGVPGVGVPGVGVPGVPGKGVPG   776

Qy    669  -GLG-----GIPPAAAAAAAKYGAAGLGVGLGGACQFPLGGVAARPGPGLSPI-PPGGACL  722
             |:|       :|     | |  :  | | |  || | |   | | | | |:  | | | | |
Db    777  PGVGVPGVGVPGVGAGSSGAGAGSSGAGAGSGAG----SGVGV-PGVGVPGVGVP   831

Qy    723  GK 724
         ||.
```

```
Db     832 GK 833
```

```
RESULT  13
US-08-642-246-15
;  Sequence 15, Application US/08642246
;  Patent No. 6033654
;  GENERAL INFORMATION:
;       APPLICANT:  SPEDRONSKY, Erwin R.
;       APPLICANT:  CAPPELLO, Joseph
;       TITLE OF INVENTION:  Tissue Adhesive Using Synthetic
;       TITLE OF INVENTION:  Crosslinking
;       NUMBER OF SEQUENCES:  35
;       CORRESPONDENCE ADDRESS:
;          ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;          STREET:  Four Embarcadero Center, Suite 200
;          CITY:  San Francisco
;          STATE:  CA
;          COUNTRY:  US
;          ZIP:  94111
;       COMPUTER READABLE FORM:
;          MEDIUM TYPE:  Floppy disk
;          COMPUTER:  IBM PC compatible
;          OPERATING SYSTEM:  PC-DOS/MS-DOS
;          SOFTWARE:  PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;          APPLICATION NUMBER:  US/08/642,246
;          FILING DATE:
;          CLASSIFICATION:  435
;       ATTORNEY/AGENT INFORMATION:
;          NAME:  ROWLAND, Bertram I
;          REGISTRATION NUMBER:  20015
;          REFERENCE/DOCKET NUMBER:  A61127-1/BIR
;       TELECOMMUNICATION INFORMATION:
;          TELEPHONE:  415-781-1989
;          TELEFAX:  415-398-3249
; INFORMATION FOR SEQ ID NO:  15:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH:  884 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;       MOLECULE TYPE:  peptide
US-08-642-246-15
```

```
Query Match          29.7%;  Score 1123;  DB 3;  Length 884;
Best Local Similarity  41.8%;  Pred. No. 6.4e-68;
Matches  352;  Conservative  53;  Mismatches  249;  Indels  188;  Gaps  56;

Qy      2  GVPG-AIPG-GVPG----GVFYPGAGLGALGGGALGPG----------GKPLKVPG-G   43
              :|||  |||  :||  |  ||  :  ||  |  :|||  |||          :|:|  :|:| |
Db     61  GVPGKGVPGVGVPGVGVGVPGVGVGVPGVGAGSSGAGAGSSGAGAGSGVGVPGVGVPGVG  120

Qy     44  LAGAGLGA-----GLGAFPAVTFPGALVPG-GVADAAAYKAAKAGAGLG------GVPGVG   93
              |||  |||:    |  ::  |  ||  |  |  ||  ||   :||:  |   |  |       ||  ||
Db    121  VPGVGVPGKGVPGVGV-PGVGVPGVGVGLPGVGVLPGGVLPGARPPGVGVLPGVPTGAGVG  179

Qy     94  --GLGVSAGAVVPQPGAGVKPGK-VPGVGLPGVTPGGVLPGARPPGVGVLPGVPTGAGVK  150
             |  ||  :  |  |  ||  |  |  ||   |||| :  |  |  ||| ||  |||| |  :  |
Db    180  VPGVGV-------PGVGV-PGKGVPGVGVPGV--GV-PGVGVPGAGAGSSGAGAGSSGAG  226

Qy    151  PKAPGVGGAFAGIPGVGPPGGPGPQPGVPLGYPIKAPKLPGGLPYTTGKLPYGYGPGPGGVA  210
              |  |  |  |  :  |   |  ||  |  |  |  :  |||:  |:  |  |  ||  |  |  | |
Db    227  AGSGAGAGSSGAGAGSSGAGAGSG---GAGAG---SGAGAGS-GVGVPGVGVPGVGVP  267

Qy    211  GAAGKAGYPTGTGVGPGQPAAAAAKAAAKFGAGAAGVLPGVG--GAGVP--GVPG-AIPG  265
              ||  | ||   | || | | |   ::: : |||:  |  |   ||  |  | |  || :| |
Db    268  GVPG--VGVP-GAGAGSSGAGAGSGAGSGVGVPGVGVPGVGVGVPGVGVPGKGVPG  325

Qy    211  GAAGKAGYPTGTGVGPGQPAAAAAKAAAKFGAGAAGVLPGVG-----GAGVP--GVPG-AIPG  265
              ||  | ||   | || | | |   ::: : |||:  |  |           ||  |  | |  |
Db    268  GVPG-VGVP-GAGAGSSGAGAGSGAGSGVGVPGVGVPGVGVGVPGVPGKGVPG  325

Qy    256  IG----GIAGVGTPAAAAAAAAAAAKAAKYGAAAGLVPGGPGPGPGVGVGAGVPGVGVP  321
              :|    |  :|  |   ||  |  |  |  |||   :   |  |:  || || |  ||| |  |
Db    326  VGVPGVGVPGVGVPGAGAGSSGAGAGS---GAGAG--SGAGAGS-GVGVPGVGVPGVP  378
```

```
Qy   322  GAGIP--VVPGAGIPGAAVPGVVSPEAAAKAAAKA-----AKYGARPGVGVG----GIPT  370
               |::|  |||:|:  |||  ::|  |||  :|:   |: ||| |||    |:|
Db   379  GVVPGKGVPGVGVPGVPGVGVPGPGAGSGSGAGSGVGVPSVGVSVP  438

Qy   371  YGVGAGGFPGFGV-GVG----GIPGVAGVPSVG-----------GVPG  402
               ||    |  | ::|  ||        |:|||:|::|           ||||
Db   439  VGVPGVGVPGKGVPGVGVPGV-GVPGAGGSGAGSGAGSGAGSGVGVPG  497

Qy   403  VG----GVPGVGISPEAQAAAARAAKY-GVGTPARAAKAAKSAAKYAAKAQ  457
               :|    |||:  |:     :||||   |||  | | :|:  |:|
Db   498  VGVPGVGVPGVGVPGVGVPGVGVPGV-GVPGAGSGCAGSGAGSGVGVPG  557

Qy   458  GVAPGVAPGVGVLAPGVGV-----GAGYPGLGV-GVGVPGLGV------GVG  620
               |  ||||| |:|:| |||       ||||| | |||:|| |:||
Db   558  GV-PGVGV-PGVGV-PGVGVPGKGVPGVGVPGVGVPGKGYPGVGVPGVGVPGVGVPGVG  716

Qy   514  LRAAAGLGAGIPGLGV-GVGPGLGV-------GAGYPGLGV-GAGVPGFGAGADEGVRRS  565
               |:|||  ||:| | ::|:||        : :||||   ||||:||
Db   613  SGAGAGSGAGSGAGSGVGVPG----------VGVPGVGVPGVPGKGVGVPGVGVPGVG  669

Qy   566  LSPELREGDPSSSQHLPSTPSSPRVPGALLAAAKAAQPGLPPTTGKLPGVGLPTGVGPGGVA  716
               |:|::| |                 |:|| || || :|:
Db   670  ------GSGAGSGAGSGAGSGVGVPG--------GVGVPGV-

Qy   621  IPG-GVVGVPGAGSGAGSGAGSGAGSGAGSGAGSGAGSGVGVPGVGVPGKGVPGVGV  776
               | |  | ||| ::|:||:|||:||:::|||
Db   717  VPGVGVPGKGVPGVGVPGVGVPGVGVPGV-GVPGAGGSGAGSGAGSGVGVPGKGVPGVGV

Qy   669  -GLG----GIPPAAAAKAAKYGAAGLGGVLGGAGQPPLGGVAARPGFGLSPI-FPGGACL  722
               | |    |     |:|    || | || |: | ||| |:|   ||  | | |||:
Db   777  PGVGVPGVGVPGAGSGAGSGAGSGAGSGAGA----SGVGV-PGVGVPGVGVGVP  831

Qy   723  GK  724
               ||
Db   832  GK  833
```

RESULT 14

PCT-US96-06229-15
; Sequence 15, Application PC/TUS9606229
; GENERAL INFORMATION:
; APPLICANT: STEDRONSKY, Erwin R.
; APPLICANT: CAPPELLO, Joseph
; TITLE OF INVENTION: Tissue Adhesive Using Synthetic
; TITLE OF INVENTION: Crosslinking
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;   STREET: Four Embarcadero Center, Suite 200
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: US
;   ZIP: 94111
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US96/06229
;   FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;   NAME: ROWLAND, Bertram I.
;   REGISTRATION NUMBER: 20015
;   REFERENCE/DOCKET NUMBER: A61127-1/BIR
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-781-1989
;   TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 884 amino acids

;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US96-06229-15

Query Match   29.7%; Score 1123; DB 5; Length 884;
Best Local Similarity 41.8%; Pred. No. 6.4e-68;
Matches 352; Conservative 53; Mismatches 249; Indels 188; Gaps 56;

```
Qy     2  GVPG-NIPG-GVVPG----GVEYPGAGLGALGGAGLGPG---------GKPLKPVPG-G  43
               |||| ||||  ||:|        ||:|:||| ||: || |:           ||   :||  |
Db    61  GVPGKGVPGVGVPGVGVPGVGVPGAGSSGAGSGAGSGAGSGAGSGVGVPGVGVPGVGVPGCVG  120

Qy    44  LAGRGLGA---GIGAFPAVTFPGALVPG-GVADAAAYKAKAGAGLG-----GVPGVG  93
               || ||:|    | ||    |||:|  ||  ||      || ||  ||       ||||||
Db   121  VPGVGVPGKGVPGVGV-PGVGVPGVPGAGSGAGSGAGSGAGSGVGVPSVG  179

Qy    94  --GLGVSAGAVVPQPGAGYKPGK-VPGVGLPGVPGGVLPGVGARFPGGVLPGVPIGCAGVK  150
               :::|||   :::|:|: |||:|   | || ||:| ||| :| ::| | |:||||:
Db   180  VPGVGV------------PGVGV-PGKGVPGVGVPGV---GV-PGVSVPGAGSGAGSGAGSGAG  226

Qy   151  FKAPGVGGAFAGIPGVGPFPGGPGPGVPLGYIPIKAPKLPGGYGLPTTGKLPTGVGPGGVA  210
               :| :| | | :|| |: | | :||| | | :| |   |:| ||:|||::| | :| :
Db   227  AGSGAGAGSGVGVPGVGVPGVPGV--GV---------GKGYP---------GVGVPGV-  267

Qy   211  GAAGKAGYPTGTGVGPGVGFPQAAAAARAAKALAAKPGAGANGVLPGVG--GAGVP--GVPG--AIPG  265
               | :||||:|| | :||||   :| ||||:||:||:|  |:|| :|| |    ||||  ||||
Db   268  GVPG-VGVP-GAGAGSGAGSGAGSGAGSGVGVPGVGVPGVGVPGCKGVPG  325

Qy   266  IG----GIAGVGVPGAAAAAAKAAKYGAAAGLVPGGPGPGFPGPGVGVPGAGVPGVGVP  321
               :|  |||| |:|||:|| |||||  |:   || ||| | |:| |||:|:|||||:||||||
Db   326  VGVPGVGVPGVGVPGAGSGSGAGS---GAGAG--SGAGAGSG-VGVPSVGVPGVGVSVP  378

Qy   322  GAGIP--VVPGAGIPGAAVPGVVSPEAAAKAAAKA----AKYGARPGVGVG----GIPT  370
               |::|  |||:|:  |||  ::|  ||| :|:  |: ||| |||    |:|
Db   379  GVVPGKGVPGVGVPGVPGVGVPGPGAGSGSGAGSGVGVPSVGVSVP  438

Qy   371  YGVGAGGFPGFGV-GVG----GIPGVAGVPSVG-----------GVPG  402
               ||    |  | ::|  ||        |:|||:|::|           ||||
Db   439  VGVPGVGVPGKGVPGVGVPGV-GVPGAGGSGAGSGAGSGAGSGVGVPG  497

Qy   403  VG----GVPGVGISPEAQAAAARAAKY-GVGTPAAAARAAKAAAQPGLVPGVGVAPGV  457
               :|    |||:  |:     :||||   |||  |   ||  ::|  ||:|||  ||||:|
Db   498  VGVPGVGVPGVGVPGVGVPGVGVPGV-GVPGAGSGCAGSGAGSGAGSGGSV  557

Qy   458  GVAPGVAPGVGVLAPGVGV----APGVGVAPGVGAPGVGAAAKSAAKVAAKAQ  513
               ||  ||||| |:|:| |||     ||| | | :| | |:||  |:|:| :|
Db   558  GV-PGVGV-PGVGV-PGVGVPGKGVPGVPGAGAGSGAGSGAGAGSGAGAG  612

Qy   514  LRAAAGLGAGIPGLGV-GVGPGLGV-------GAGVPGLGV-GAGVPGFGAGADEGVRRS  565
               |:|||  ||:| | ::|:||        : :||||   ||||:||  |::|:|::|
Db   613  SGAGAGSGVGVPGVGVPGVPGVGVPGKGVGVPGVGVPGVPGGVPGVPGAGAGSGAGA---  669

Qy   566  LSPELREGDPSSSQHLPSTPSSPRVPGALLAAAKAKAAKYGAAVPGV-LGGLGALG----GVG  620
               |:|::| |                 : :|: :|| | ||::|
Db   670  ------GSGAGSGAGSGAGSGVGVPG------VGVPGVGVPGVGVPGKGVPGVGVPGVGV  716

Qy   621  IPG-GVVGAPAAAAAAKAAKAAQPGLVGAAGLGGLGLGVGGLGVPGVG---------------  668
               | |  | ||| :|:||:|::|:|||: :|| |::||||:|||:|||
Db   717  VPGVGVPGKGVPGVGVPGVGVPGVGVPGKGVPGVGVPGVGVPGVGVGVG  776

Qy   669  -GLG----GIPPAAAAKAAKYGAAGLGGVLGGAGQPPLGGVAARPGFGLSPI-FPGGACL  722
               | |    |     |:|    || | || |: | ||| |:|   ||  | | |||:
Db   777  VPGVGVPGVGVPGAGSGAGSGAGSGAGSGAGA----SGVGV-PGVGVPGVGVGVP  831

Qy   723  GK  724
               ||
Db   832  GK  833
```

RESULT 15

```
US-08-707-237A-103
; Sequence 103, Application US/08707237A
; Patent No. 5830713
; GENERAL INFORMATION:
;    APPLICANT: Ferrari, Franco A.
;    APPLICANT: Capello, Joseph
;    APPLICANT: Crissman, John W.
;    APPLICANT: Dorman, Mary A.
;    TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC
;    TITLE OF INVENTION: REPETITIVE DNA
;    NUMBER OF SEQUENCES: 108
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;       STREET: Four Embarcadero Center, Suite 3400
;       CITY: San Francisco
;       STATE: California
;       COUNTRY: United States
;       ZIP: 94111-4187
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/707,237A
;       FILING DATE: 03-SEP-1996
;       CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/175,155
;       FILING DATE: 29-DEC-1993
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/053,049
;       FILING DATE: 22-APR-1993
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 07/609,716
;       FILING DATE: 06-NOV-1990
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 07/269,429
;       FILING DATE: 09-NOV-1988
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 07/114,618
;       FILING DATE: 29-OCT-1987
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 06/927,258
;       FILING DATE: 04-NOV-1986
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Trecartin, Richard F.
;       REGISTRATION NUMBER: 31,801
;       REFERENCE/DOCKET NUMBER: A-55186-10/WHD
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (415) 781-1989
;       TELEFAX: (415) 398-3249
;       TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 103:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH: 1002 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: unknown
;       TOPOLOGY: unknown
;    MOLECULE TYPE: protein
US-08-707-237A-103

Query Match      29.6%;  Score 1122;  DB 2;  Length 1002;
Best Local Similarity  39.1%;  Pred. No. 8.5e-68;
Matches 370;  Conservative  54;  Mismatches 271;  Indels 252;  Gaps 60;

Qy   2 GVPG-AIPG-GVPG--------GVPYPGAGLGALGGGALGPG-------------GKPLKPVP 41
         |||| |||| ||||        || ||||| | ||||  ||             | ||||
Db  63 GVPGKGVPGVGVPGVGVPGVGVPGVPGAGAGSGAGSGAGSSGVGPGVGVGVPGVGVPGKGVP 122

Qy  42 G---------GLAGAGLGAGLGAPPAVTPPGALVPG-GVADAAAAYKAAKAGAGLG-- 87
         |          |: ||| |:|||  | |: ||| ||
```

```
Db  123 GVGVPGVGVPGVGVPGVGAGSGAGSGAGSSGVGPGVGVGVPGVGVPGKGV-PGVGVPGV 181
                                                                       ||||

Qy   88 GVPGVG-------GLGVSAGAVVPQPGAGVKPGKVPGVGLPGV-TPGGVLPGARPPGVGVL 140
           |||||        | |:|  || ||| | | ||||:|   |||| :|| ||||||||||

Db  182 GVPGVSVPGAGSGAGSGAGSSGVGPGVGVGVPGVGVPGVGVPGVGVPGVGVGVPGVGVPGV- 240
                                                              .||||

Qy  141 PGVPGAGVKPKA----PGVGGAAPAGIPGVGPPGGPQPGVPGLGVP-IKAPKLPGGYGLPY 195
          || ||: |:|    |||||| |   | :|||| |||  |  ||   | |: ||||||| |

Db  241 PGAGAGSGAGSGAGSGVGPGVGVPGVGVPGKGVPGV---GVPGVGVP----GVGVP- 293
          |||||||||||||||||||||                                |:|

Qy  196 TTGKLPYGYPGGVAGAAGKAGP---------------------TGT-GVGPQAAAAA 233
           |:| || ||||||||||| |                     ||| ||||||||

Db  294 G------GAGAGSGAGSGAGSGVGVPGVGVGVGVPGKGVPGVGVGVGVGVPGVPGVPG 346
           :|| |||| | ||||  ||||| || |||||||| ||  |:| |||

Qy  234 AKAAKFGAGAAGVLPGVG---GAGVP-GVPG-AIPGIG----GIAGVGTPAAAAAAAA 284
          |: | |||: :|| | :||||   |||| ||||| :|   |: ||||||:::: :

Db  347 AGAGSGAGAGSGVGPGVGVPGVGVPGVPGKGVPGVGVPGVGVGVPGVPGAGAGSGAG 405
                                                                   : |

Qy  285 AKAAKYGAAAGL----------------VPGG---------PGVG-PGV------ 307
          |: |:  ||  |                 ||:           ||:   ||:

Db  407 A---------GSGVGVPGVGVGVGVPGVGVPGVGVPGVGVPGVPGAGAGSGGAGSGAGS 461
          |:

Qy  308 -VGVPGAGVPGVGVPGAGIP---VVPGAGIPGAAVPGVVSPEAAAKAAAKAAAKIYGARPGVG 364
          || |||| ||||||||  |    ||||| :||| |:||| | |:||||||||| |||:||||

Db  462 GVGVPGVGVPGVGVPGVGVPGKGVPGVGVPGVGVPGAGAGSGAGSGAGSGVGPGVVG 521
                                                                   |||

Qy  365 VG--GIPTYGVGAGGFPGPGVGVGGIPGVAGVPGVG----------GVPGVG-----GYP 407
          |:  |:  ||:|  | |||| || |:|||| |: || |           |||||   |||

Db  522 VPGVGVPGVGVPGKGVPGVGVGVPGVGV-GVPGAGAGSGAGSGAGSGVGPGVGVPGVGVGVP 580
          |||||||||||||||||||||

Qy  408 GVGISPEAQAAAAAKAAKY-GVGTPAAAAAKAAAKAAQPGLVPGVGVAPGVGVAPGVGV- 465
          || :|| ||||||||:| |||:| ||||||:|||  | || ||||| |:|||||| |||

Db  581 GVGVPGVGVPGVGVPGVGVPGVGVPGAGSGAGSGAGSGVG-VPGVGV-PGVGV-PGVGVP 637
          |:|  ||

Qy  466 ---APGVGLAPGVGVAPGVGV-------------------APGVGVAPGIGPGGVAAAAKSAA 506
             |||  | ||||| ||| |                   |||| | | |:|||||||:||

Db  638 GKGVPGVGV-PGVGV-PGVGVPGVGVPGAGAGSSGAGSGAGSGVGVCVPGVGV-PGVGVPGKGVP 694
          ||: |||   |:| |

Qy  507 KV-----------AAKAQLRAAAGLGAGIPGLGV-GVGVPGLGV-----GAGVPGLG 546
          |:            || || |:|| ||||| :||| |||||||  |     ||||||||

Db  695 GVGVPGVGVPGVGVPGVGAGSGAGSGAGSSGVGPGVGVGVPGVGVPGKGVPGVGVPGVG 754
                                                                   ||||

Qy  547 V-GAGVPGFGAGADFGVTRRSLS------------PELREGDFSSSQHLPSTPS----SPRV 590
          ||||| | ||||  | ||||:|           |  || ||:| |||| :||   |||:

Db  755 VPGVGVPGAGAGSGAGSGAGSGVGPGVGVPGVGVPGVGV---GVPGKGVPGVGVGVPGVGV 812
          ||||||||

Qy  591 PGALAAAAA-AKYGAAVPGV------LGGLGALG---GVGLPG-GVVGAG-PAAAAAAA 637
          |||| ||: ::||| ||||       |||||:|    |||||: |: |||: |||||||

Db  813 PGAGAGSGAGSGAGSGVGPGVGVPGVGVPGKGVPGVGVGVPGVGVPGVGVPGAGAGSG 872
                                                                   |||

Qy  638 KAAAKAAQPGLVGRAAGLG--GLGVGGLGVPGVG--GLG----GIPPAAAAAKARYGA-AG 688
          |||| | || |||:|| |   ||||||| | |:| |||     |:|||||||:|||||: ||

Db  873 AGAGSGVGVPGVGVPGVGVPGVGKGVGVPGVGVPGVGVPGVGVPGVPGAGAGSGAGSGVG 932
                                                                   ||||

Qy  689 LGGV-LGGAGQPGLGGVAAR------PGFGLSPIFPGGACLGKACG 727
          |||  | |:|||| | || :||       |: :|: || |::|  |:

Db  933 VPGVGVPGVG--VPGVGVPGKGVPGVGVPGVGVPGVPGAGAGSGAG 978
          || |:  |:: :| | || | || |
```

Search completed: April 24, 2001, 16:36:20
Job time: 300 sec

```
            GenCore version 4.5
       Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    April 24, 2001, 16:34:15 ; Search time 74.56 Seconds
                                       (without alignments)
                                       673.772 Million cell updates/sec

Title:     US-09-340-736-1
Perfect score: 3785
Sequence:  1 GGVPGAIPGGVPGGVPYPGA.......LSPIPPGACLGKACGRKRK 731

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

rched:     198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters:   198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR_67:*
        1:   pir1:*
        2:   pir2:*
        3:   pir3:*
        4:   pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3747.5	99.0	792	1	EAHU	elastin precursor,
2	2603.5	68.8	770	2	S59623	tropoelastin - she
3	2488	65.7	747	1	EABO	elastin precursor,
4	2387	63.1	860	1	EAMS	elastin precursor,
5	2311	61.1	864	1	EANT	elastin precursor
6	1696.5	44.8	784	2	A26601	elastin precursor
7	575.5	17.8	1758	2	T29350	hypothetical prote
8	575.5	17.8	1759	2	T29351	collagen alpha 2(I
9	663	17.5	1901	2	F70806	hypothetical glyci
10	661.5	17.5	718	2	A36068	major ampullate fi
11	644	17.0	1763	2	S16366	collagen alpha 2(I
12	635	16.8	2639	2	T31238	fibroin - Chinese
13	626	16.5	1489	1	D70807	hypothetical glyci
14	611	16.1	749	2	A70812	hypothetical glyci
15	609.5	16.1	1669	1	CGHU4B	collagen alpha 1(I
16	599.5	15.8	1691	1	S22917	collagen alpha 5(I
17	595.5	15.7	1306	2	A70034	hypothetical glyci
18	595.5	15.7	882	2	B70812	hypothetical glyci
19	591	15.6	1660	2	A70809	hypothetical glyci
20	589.5	15.6	1669	1	CGMS4B	collagen alpha 1(I
21	583	15.4	1329	2	E70937	hypothetical glyci
22	582	15.4	1464	1	CGHU1S	collagen alpha 1(I
23	580	15.3	778	2	F70963	hypothetical glyci
24	577.5	15.3	1042	1	CGCH1S	collagen alpha 1(I
25	575.5	15.2	1466	1	CGHU7L	collagen alpha 1(I
26	573	15.1	767	2	E70895	hypothetical glyci
27	570	15.1	1049	1	CGBO7S	collagen alpha 1(I
28	567	15.0	783	2	E70824	hypothetical glyci
29	567	15.0	1418	2	T45467	collagen alpha 1(I

30	566.5	15.0	1691	1	CGHU6B	collagen alpha 6(I
31	564.5	14.9	812	2	S31521	collagen COLF1 - f
32	563	14.9	1487	1	CGHU6C	collagen alpha 1(I
33	560.5	14.8	1453	1	S21626	collagen alpha 1(I
34	560	14.8	1464	1	S59856	collagen alpha 1(I
35	559.5	14.8	1079	2	B70807	hypothetical glyci
36	559	14.8	1373	1	A43291	collagen alpha 2(I
37	558	14.7	1419	2	A41182	collagen alpha 1(I
38	558	14.7	1487	2	B41182	collagen alpha 1(I
39	556.5	14.7	1366	1	CGHU2S	collagen alpha 2(I
40	553	14.6	1712	1	CGHU2B	collagen alpha 2(I
41	551	14.6	957	1	D70835	hypothetical glyci
42	549	14.5	1381	1	E70806	hypothetical glyci
43	544	14.4	853	2	A70896	hypothetical glyci
44	544	14.4	1497	1	I49607	procollagen type V
45	543.5	14.4	1492	2	A40333	collagen alpha 1'(

ALIGNMENTS

RESULT   1
EAHU
elastin precursor, long splice form - human
N;Alternate names: tropoelastin
C;Species: Homo sapiens (man)
C;Date: 22-Jun-1990 #sequence_revision 26-Jul-1996 #text_change 22-Jun-1999
C;Accession: A32707; A33705; A30524; A53891
R;Indik, Z.; Yeh, H.; Ornstein-Goldstein, N.; Sheppard, P.; Anderson, N.; Rosenbloom,
Proc. Natl. Acad. Sci. U.S.A. 84, 5680-5684, 1987
A;Title: Alternative splicing of human elastin mRNA indicated by sequence analysis of
A;Reference number: A32707; MUID:87289668
A;Accession: A32707
A;Molecule type: mRNA
A;Residues: 1-500,507-792 <IND>
A;Cross-references: G3:M16983; GB:J02948
R;Bashir, M.M.; Indik, Z.; Yeh, H.; Ornstein-Goldstein, N.; Rosenbloom, J.C.; Abrams,
J.Biol. Chem. 264, 8887-8891, 1989
A;Title: Characterization of the complete human elastin gene. Delineation of unusual
A;Reference number: R33705; MUID:89255358
A;Accession: A33705
A;Molecule type: DNA
A;Residues: 1-27 <BAS>
A;Cross-references: G3:J04821; NID:g182052; PIDN:AAA52379.1; PID:g553276
R;Fazio, M.J.; Olsen, D.R.; Kauh, E.A.; Baldwin, C.T.; Indik, Z.; Ornstein-Goldstein,
J. Invest. Dermatol. 91, 458-464, 1988
A;Title: Cloning of full-length elastin cDNAs from a human skin fibroblast recombinan
A;Reference number: A30524; MUID:89009960
A;Accession: A30524
A;Molecule type: mRNA
A;Residues: 1-453,483-617,651-792 <FAS>
A;Cross-references: EMBL:M36860; NID:g182061; PIDN:AAA52382.1; PID:g182062
A;Note: this sequence represents a composite of several splice forms
R;Fazio, M.J.; Olsen, D.R.; Kuivaniemi, H.; Chu, M.L.; Davidson, J.M.; Rosenbloom, J.
Lab. Invest. 58, 270-277, 1988
A;Title: Isolation and characterization of human elastin cDNAs, and age-associated va
A;Reference number: A53891; MUID:88156138
A;Accession: A53891
A;Molecule type: mRNA
A;Residues: 1-453,483-500,507-617,651-792 <FA2>
A;Cross-references: G3:M24782; NID:g182063; PIDN:AAA53190.1; PID:g182064
A;Comment: The term tropoelastin refers to a soluble precursor form of the extracellu
ine oxidase activity.
C;Genetics:
A;Gene: GDB:ELN
A;Cross-references: GDB:119107; OMIM:130160
A;Map position: 7q11.23-7q11.23
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-792/Product: elastin #status predicted <MAT>
F;782-787/Disulfide bonds: #status predicted

Query Match 99.0%; Score 3747.5; DB 1; Length 792;
Best Local Similarity 95.4%; Pred. No. 8.3e-166;
Matches 731; Conservative 0; Mismatches 0; Indels 35; Gaps 2;

```
Qy    1 GGVPGAIPGGVPGGVFYPGAGLGALGGGAIPGGGKPLKPVPGGLGAGLGAGLGARPAVT  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27 GGVPGAIPGGVPGGVFYPGAGLGALGGGAIPGGGALPGGGKPLKPVPGGLGAGLGAGLGARPAVT  86

Qy   61 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPGPGAGVKPGKVPGGVGL  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPGPGAGVKPGKVPGGVGL  146

Qy  121 PGVYPGGVLPGARFPGGVGLPGVPTGAGVRKPAPGVGGAFAGIPGVGPFGGPGPGVPLGY  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  147 PGVYPGGVLPGARFPGGVGLPGVPTGAGVKPKAPGVGGAFAGIPGVGPFGGPGPGVPLGY  206

Qy  181 PIKAPKLPGGYGLPYYTGKLPYYTGKLPGGYGVGPGVAGGAGKAGYPTGTGVGPQAAAAAKAAAKP  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  207 PIKAPKLPGGYGLPYYTGKLPYYTGKLPGGYGVGPGVAGGAGKAGYPTGTGVGPQAAAAAKAAAKP  266

Qy  241 GAGAAGVLPGVGGAGYPGVPGAIPGIGGIAGVTPAAAAAAAAKAAAKYGAAAGLVPGG  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  267 GAGAAGVLPGVGGAGYPGVPGAIPGIGGIAGVTPAAAAAAAAKAAAKYGAAAGLVPGG  326

Qy  301 PGTGPGVVGVGVPGVSAGVPGVGVPGVSAGIPVPGAGIPGAAVPGVVSPEAAAKAAAKAAAKYGAR  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  327 PGTGPGVVGVGVPGVSAGVPGVGVPGVSAGIPVPGAGIPGAAVPGVVSPEAAAKAAAKAAAKYGAR  386

Qy  361 PGVGVGGIPTYGVGAGGFPGFGVGVGGIPGVAGVSVGGIPGVAGVSVGGVPGVGVPGVPGVGG  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  387 PGVGVGGIPTYGVGAGGFPGFGVGVGGIPGVAGVSVGGIPGVAGVSVGGVPGVGVPGVPGVGG  446

Qy  421 AKAAKY------------------SVGTPAAAAAKAAAKAAQF------  445
         |||||                   |||||||||||||||||||
Db  447 AKAAKYGAAGAGVLGGLVPGPDAAVPGVSVTGGVPGVSVPGVGVARPGVGVAPGVSVPGPAALLNLA  506

Qy  446 GLVPGCVTAPGVGVAPGVGVARPGVGVAPGVGVARPGVGVAPGVGVAGGVPGVGVGGGVPGVGGVPGVGGVPGVGGV  505
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  507 GLVPGCVTAPGVGVAPGVGVARPGVGVAPGVGVARPGVGVAPGVGVAGGVPGVGVGGGVPGVGGVPGVGGVPGVGGV  566

Qy  506 AKVAAKAQLRAAAGLGAGPPLGGVARPGFGLSPIFPGGACLGKACGKKKK  555
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  567 AKVAAKAQLRAAAGLGAGPPLGGVARPGFGLSPIFPGGACLGKACGKKKK  626

Qy  566 LSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGVLGGLGALGGVGIPGGV  625
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  627 LSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGVLGGLGALGGVGIPGGV  686

Qy  626 VGAGPAAAAAAKAAAAKAAAAKAAQFGLVGAAGLGGLGVGVGGGLGVGGLGVGGLGVGGLGIPPAAAAAAKAKYG  685
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  687 VGAGPAAAAAAKAAAAKAAAAKAAQFGLVGAAGLGGLGVGVGGGLGVGGLGVGGLGVGGLGIPPAAAAAAKAKYG  746

Qy  686 AAGLGGVLGGAGQPPLGGVAARPGFGLSPIFPGGACLGKACGRKKK  731
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  747 AAGLGGVLGGAGQPPLGGVAARPGFGLSPIFPGGACLGKACGRKKK  792
```

RESULT 2
S59623
tropoelastin - sheep
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 23-Aug-1996 #sequence_revision 13-Mar-1997 #text_change 24-Oct-1997
C;Accession: S59623; A24758
R;Mauch, J.C.; Sandberg, L.B.; Roos, P.J.; Jimenez, F.; Christiano, A.M.; Deak, S.B.; Bc
Matrix Biol. 14, 635-641, 1994
A;Title: Extensive alternate exon usage at the 5' end of the sheep tropoelastin gene.
A;Reference number: S59623
A;Accession: S59623
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-770 <MAU>
R;Yoon, K.; Davidson, J.M.; Boyd, C.; May, M.; LuValle, P.; Ornstein-Goldstein, N.; Smit

---

Arch. Biochem. Biophys. 241, 684-691, 1985
A;Title: Analysis of the 3' region of the sheep elastin gene.
A;Reference number: A24758; MUID:85305763
A;Accession: A24758
A;Molecule type: mRNA
A;Residues: 655-669,671-716,732-770 <YOO>
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;760-765/Disulfide bonds: #status predicted

Query Match 68.8%; Score 2603.5; DB 2; Length 770;
Best Local Similarity 67.9%; Pred. No. 3.1e-113;
Matches 553; Conservative 26; Mismatches 81; Indels 155; Gaps 21;

```
Qy    1 GGVPGAIPGGVPGGVFYPGAGLGALGGGALPGGGKPLKPVPGGLAGAGLGA-GLGAPPAV  59
         ||| ||:|||||||||||||||||||| |: ||||||| ||| ||| | || |  | |
Db   27 GGVLGAVPGGVPGGVFPPGAGLGLGVGLGAPGVKPAKPGVGGLVPGGLGAEGSGLLPAG  86

Qy   60 TPFGAL--VPGGVLDAAAAYK-AAKAGA---GLGGVPGVGGLGVSAGAVVPQPGA----  108
         ||::||  | ||||| :| |||: |:|:||     |  ||| || || || |||  ||
Db   87 AFPGGFPGAGGLAGVPGVLPGVTPGVGVLP--GARFPGVGLPGVPTGAGVKPKAPGVGGLPGV  146

Qy  109 GYKPGKVPGVGLPGVYPGGVLP--GARFPGVGVLPGVPGGVLPCTGVLPYTGKLPYTGK  166
         |:|||:|||:|||||||||||   || |||||: ||| |: || |||||: |:|:||:|
Db  147 GVKPGKVPGVGLPGVYPGGVLPGTGKLPCTGVLPCTGVLPYTGKLPYTGKLPGSGLPGGV  206

Qy  167 GPPGGQPGVPLGYPIKAPKLPGGYGLPYTGKLPYGVPGVGPGVAGAAGKAGYPTGTCVGP  226
         |||||:||||||||||||||||||||||: |:|| ||| || ||||||||||||||||||
Db  207 GPPGGQPGVPLGYPIKAPKLPGGYGLPYSTGKLPYGYGPGDVAGAAGKAGYPTGTCVGP  266

Qy  227 QAAAAAAKAAAKPGAGAAGVLP--GVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAA  284
         |||||||||||||| |         |  || ||||  || ||:||: |||||||||||
Db  267 Q-AAAAAAKAAAKLGAGGAGVLPGVGVGGAGIPGAGAIPGIGGIAGAGAPDAAAAAAA  325

Qy  285 AKAAKYGAAAGLVPGGPGFGPGVGVYVPGAGYPVPGAGIP---------------V  327
         |||||||  |   |                     ||| |: :
Db  326 AKAAKYGAAG-------VPGPGVGVPGAGVPGVGVPGVGVPGVGVPGVGVG  378

Qy  328 VPGAGIPGAANPGVVSPEAAAKAAAKAAKYGARPGVVGGIPTYGVAGGPGFG-----  382
         ||||||||||  || |||:||||||||||:|||| || || ||| |: |:|
Db  379 VPGVLPGVVPGAVSPAAAAAKAAAKAAKPGARGGVVGGIPTFGVGPGDFGPGFGDAAAQ  438

Qy  383 --------YGVGGIPGVAG-VP----SVGGVPGVGVPGVGISPEAQAAAAAKAAKYGV  428
                   ::|| |  || |:       :||||||||||| || ||| ||| |
Db  439 AAAAKAAAKLGAGGVGVLPGVGPDAAVPGFGLVPGVGVAPGVGVAPGVGVAPGVGVAPGVGVG  479

Qy  429 GTPAAAAKAAAKAAQFGLVPGVGVAPGVGVARPGVGVAPGVGVAPGVGVAPGVGVAPGI-  488
         ||:||||||||||||| |: || || ||| || || || || || ||||
Db  480 GAPAAAAAAKAAKAAQFGLGPGVGVAPGVGVARPGVGVAPGVGVAPGVGVARPGI-  538

Qy  489 VAPGIPGGV--AAAKSAAKYAAKAQLRAAAGLGAGIPGLGVGVGVGAGVPGLG  546
         || | |||   :::|| |||: || |:|||||||||:||| |||| ||| |||
Db  539 --GLPGPGVIAPARAKSAAKAAAKAAQPRAALGIPAGVPGLGVGVGVPGLGVGVPGLG  595

Qy  547 VGAGVPGFGAGADEGVERSLSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAV  606
         ||| || || :   :: :  :   ||||||||||||||||||||| |||||||||
Db  596 AGA---------------------------------------VPGTLAAAKAAKPAPGG  615

Qy  607 PGVLGGLGALGGVGIPGGVVGAGPAAAAAKAAAKAAAKAAAQFGL--VGAALGLGLG-VGGLG  663
         ||||||||| || | ||||| | |||||||||||||| |||||:||: ||:|||:|| |||:|
Db  616 VGALGGVGDLGGAGIPGVGGGVPAAAALGAKAAAKAAAKAAQFGLGGLGVGGLGVGGLGAVGGLG  675

Qy  664 -VPGVGVGGLGGTIPPAAAAKAAKYGAAGLGAGIPGLGAGQFPLGGVAARPGFGLSPI------  715
          ||||| ||  |||  |||||| |||| | |||||||||| || |||||| |||||
Db  676 AVPGAVGLGGVSPAAAAAKAAKFGAGLGVLGAGRPFFGGVLGARPFFGGVLGLSPIFPGGAGG  735

Qy  716 ----------------------FPGGACLGKACGRKKK  731
                               ||||||||||||| ||
Db  736 LGVGGKPKPFGPGALGALGFPGGACLGKACGSCGRKKK  770
```

RESULT 3

BABO
elastin precursor, splice form a - bovine
N;Alternate names: tropoelastin
N;Contains: elastin precursor splice form b; elastin precursor, splice form c
C;Species: Bos primigenius taurus (cattle)
C;Date: 08-Jun-1989 #sequence_revision 26-Jul-1996 #text_change 22-Jun-1999
C;Accession: A31865; A26728; B26728; C26728; A22343; I45886
R;Yeh, H.; Anderson, N.; Ornstein-Goldstein, N.; Bashir, M.M.; Rosenbloom, J.C.; Abrams,
Biochemistry 28, 2365-2370, 1989
A;Title: Structure of the bovine elastin gene and S1 nuclease analysis of alternative sp
A;Reference number: A31865; MUID:89271459
A;Accession: A31865
A;Molecule type: DNA
A;Residues: 1-27 <YEH>
A;Cross-references: GB;J02855; NID:g340504; PIDN:AAA30776.1; PID:g552339
R;Raju, K.; Anwar, R.A.
J. Biol. Chem. 262, 5755-5762, 1987
A;Title: Primary structures of bovine elastin a, b, and c deduced from the sequences of
reference number: A92640; MUID:87194772
Accession: A26728
A;Molecule type: mRNA
A;Residues: 1;'RS',4-11,'E',13-636,'V',638-747 <RA1>
A;Cross-references: GB;K02717; NID:g163019; PIDN:AAA30503.1; PID:g163020
A;Accession: B26728
A;Molecule type: mRNA
A;Residues: 1,'RS',4-11,'E',13-225,240-636,'V',638-747 <RA2>
A;Cross-references: GB;K03505; NID:g163025; PIDN:AAA30505.1; PID:g163026
A;Accession: C26728
A;Molecule type: mRNA
A;Residues: 1,'RS',4-11,'E',13-225,260-636,'V',638-747 <RA3>
A;Cross-references: GB;K03506; NID:g163027; PID:g163028
Biochemistry 24, 3075-3080, 1985
A;Title: Structure of the 3' portion of the bovine elastin gene.
A;Reference number: A22343; MUID:85280426
A;Accession: A22343
A;Molecule type: DNA
A;Residues: 613-747 <CIC>
A;Cross-references: GB;M20415
R;Rosenbloom, J.
Lab. Invest. 51, 605-623, 1984
A;Title: Biology of disease: Elastin: Relation of protein and gene structure to disease.
A;Reference number: I45886; MUID:85059254
A;Accession: I45886
A;Molecule type: DNA
A;Residues: preliminary; translated from GB/EMBL/DDBJ
sidues: 678-683,685-747 <ROS>
cross-references: GB;M31898; NID:g163015, PIDN:AAA96417.1; PID:g163018
R;Brown, P.L.; Mecham, L.; Tisdale, C.; Mechan. R.P.
Biochem. Biophys. Res. Commun. 186, 549-555, 1992
A;Title: The cysteine residues in the carboxy terminal domain of tropoelastin form an in
A;Reference number: A58621; MUID:92337651
A;Contents: annotation, disulfide bonds
C;Comment: The term tropoelastin refers to a soluble precursor form of the extracellular
ine oxidase activity.
C;Genetics:
A;Introns: 634/3; 653/3; 676/3; 689/3; 707/3; 716/3; 733/3
A;Note: the list of introns is incomplete
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;1-747/Product: elastin precursor, splice form a #status predicted <PPA>
F;1-225,260-747/Product: elastin precursor, splice form c #status predicted <PPC>
F;1-225,240-747/Product: elastin precursor, splice form b #status predicted <PPB>
F;1-25/Domain: signal sequence #status predicted <SIG>
F;27-747/Product: elastin #status predicted <MAT>
F;105,109,252,271,275,324,327,400,404,407,445,448,489,493,544,544,548,552,606,609,645,649,68
F;737-742/Disulfide bonds: #status experimental

Query Match 65.7%; Score 2488; DB 1; Length 747;
Best Local Similarity 67.0%; Pred. No. 6.1e-108;
Matches 529; Conservative 35; Mismatches 98; Indels 98; Gaps 22;

```
Qy    1 GGYPGAIPGGVPGGVPTYPGAGLGALGGALGPGGKPLKPVPGGLAGGAGGLGAGLGAPPAVT  60
          |||:||:||||||||| :|||||| :|||| || |  ||| :|| || |||| | ||||
Db   27 GGVPGAVPGGVPGGVFFPGAGLGGLGVGGLGPGVKPAKPGVGGLVPGGLGAEGSALPG-A   85

Qy   51 PGAL--VEGGVADAAAYK-AAKAGA---GLGGVPGGVGLGVSAGAVPVPQ----PGAGV  110
          |||  |:|| | :||| | ||||:| | |||:||||| |||||| |:||      ||||
Db   86 PPGGFFGAGGSAAGAAGAAAYKAAAKAGAAGLGVGGLGGVGGLGVSTGAVVPGLGAGVGAGV  145

Qy  111 KPGKVPGVGLPGVYPGGVLP--GAKFPGVGLVPGTPGAGVKPKAPGVGGAFGAGIPVGGP  168
          |:||||||||||||||||||  ||:|||:|:|||:||:||||||| | |||||||||||||
Db  146 KEGKVPGVGLPGVYPGGVLPGAGARFPGIGVLPGVPGTPGAGVKPKAQVGAGAFGAGIPVGGP  205

Qy  169 FGGPQPGVPLGYPIKAPKLPGGYGLPYTTGKLPYGYGPGGVAGGAAGKAGTPTGTGVGPQA  228
          |||:|||:|||||||||||| ||||||:|||||||:|||||||||| ||||||||||||| |
Db  206 FGGQQPGLPLGYPIKAPKLPAGYGLPYKTGKLPYGFGPGGVAGGAGKAGTPTGTGVGPQ-  264

Qy  229 AAAAAKAAAKFGAGAGVLP--GVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAAK  286
          |||||:||||:||| ||||  ||| || ||||:||||||||||||| | |||||||||||
Db  265 AAAAAAKAAAKLGAGGAGVLPGVGVGGPGIFGAPGAIPGIGGIAGVGAPDAAAAAAAAK  324

Qy  287 AAKYGAAAGL----VP--GGPGPG-PGV----VGVPGAGVPGVGYPGAGIP--VVPGAGI  333
          |||:|||:|:|    :|  |||:|| |||    ||||||:|||||||| :|  |||||||
Db  325 AAKFGAAGGLGPGVGVGVGGPVGVGGVGVPGGVGVPGSVGVPGVGYPGGVPGGVGV  384

Qy  334 PGAAVPGVVSPEAAAKAAAKAAKYGARPGVGVGGIPTYGVGAGGPPG---------  382
          ||:|:||| ||||:||||:|||:|||:|| | ||||||:|| |:| ||
Db  385 PGVGVPGALSPAAPAKAAKAAKFGARGAVGIGGIPTPGLGPGPGIGDAAAAPAAAAA  444

Qy  383 ---VGVGGIPGVAGV-----PSVGGVPGVGGVPGVGISPEAQARAAAAAAKYGVGTPAA  433
             |||:| :||:|:|     :||||||||||:|||| |:||| ||||||||:|||||||
Db  445 KAAKIGAGGVGALGGGVTPGGAPGAIPGLPGVGGVPGV---------------PAA  485

Qy  434 AAAKAAAAAQFGLVPGVGVAPGVGVAPGVGVAPGVGVAPGVGVGVAPGI  493
          ||||||||||||||||||||||||||||||||||:| :|:  |||  |
Db  486 AAAKAAAAAAQFGLAPGVGVAPGVGVTPGVGVPGGVVPGVVPGIGLGPG---------GV  534

Qy  494 GPGGVAAAAKSAAAKVAAKAQLRAAAGISGAGIPGLGVGVGVGLGVGAGVPG  553
          ||| ||||||||| |||||||:||||||:| ||||||||||||||||||||
Db  535 IGAGVPAAAKSAAKAAAKAQFRAAAGLPAGVPGLGVGAGVPGLGVGVGAGVPG  594

Qy  554 FGAGADEGVKRSLLSPELREGDFSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGVLGGL  613
          ||                           | :||||||||||||:|:|||:||| |||||:|
Db  595 PGA-----------------------------VPGTLAAAKRAAKFGPGGVGALGGV  621

Qy  614 GALGGVGIPGGVVGAGPAAAAAAAKAAAAKAAQFGLVGAAGLAGLGVVGGLG-VPGVGGLGG  672
          | |||||||||| |:| |||||:||||:|||||||:|   || |||| |||  ||| ||||
Db  622 GDLGGAGIPGGVAGVP-AAAAAKAAAKAAQFGSL---GGVGGLGVGGLGAVPGAVGLGG  677

Qy  673 IPPAAAAKAAKYGAAGLGGVLGAGQPLGVAARPGFGLSPI----------FPGGAC  721
          : |||||||||:|||||||||||||| | | ||:|||||            |||||
Db  678 VSPAAAAKAAKFGAAGLGGVLGAGQPFPIGGGAGGLGVGGKPFKPFGGALGALGFPGGAC  737

Qy  722 LGKACGRKRK  731
          ||| ||||||
Db  738 LGKSCGRKRK  747
```

RESULT 4
EAMS
elastin precursor - mouse
N;Alternate names: tropoelastin
C;Species: Mus musculus (house mouse)
C;Date: 18-Aug-1995 #sequence_revision 16-Aug-1996 #text_change 22-Jun-1999
C;Accession: A57721
R;Wydner, K.S.; Sechler, J.L.; Boyd, C.D.; Passmore, H.C.
Genomics 23, 125-131, 1994
A;Title: Use of an intron length polymorphism to localize the tropoelastin gene to mo
A;Reference number: A57721; MUID:95130069
A;Accession: A57721
A;Molecule type: mRNA
A;Residues: 1-860 <WYD>
A;Cross-references: GB;U08210; NID:g473273; PIDN:AAA80155.1; PID:g473274
C;Genetics:

A;Map position: 5
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-860/Product: elastin #status predicted <MAT>
F;850-855/Disulfide bonds: #status predicted

Query Match   63.1%;   Score 2337;   DB 1;   Length 860;
Best Local Similarity   61.1%;   Pred. No. 3e-103;
Matches 546;   Conservative   25;   Mismatches   96;   Indels 226;   Gaps   34;

```
Qy    1  GGVPGAIPGGSVPGGVFYPGAGLGAL--GGGALGPGGKPLKP--------VPGGLAGAG   48
Db   32  GAVPGGLPGGVPGGYYPGAGIGGLGGGGALGPGGPKPKPGAGLLJTPGAGPGGLGGAG   91

Qy   49  LGAGGLGAFPAVTFP--GALVPGGVADAAAAYK-AAKAGAGLGGVPGV----------   92
Db   92  PGAGLGAFPAGTFPGGAGALVPGGAGALVPGGAGAVPGGVGGVGGVPGGVGGVGGVPGGV  151

Qy   93  ----------------GGLGVSAGAVPQPGAGV---XPGKVPGVGLPGVV          124
Db  152  GVGGVPGGVGVGGVGIGGLGVSTGAVVPQVQVAGVGIGAGGKPGGKVPGVGLPGVV     211

Qy  125  PGGVLP--GARFPGVGVLPGVFPGAGVKPAFGVPGGAFAGIPGVGFGGPGPGVPLGYPI  182
Db  212  PGGVLPGTGARFPGVGVLPGVPFGTGVKAKAPGGGGAPSGIPGVGVPGGVQPGGVPLGYPI  271

Qy  183  KAPKLPGGYGLPVTTGKLPYGYGPGGVAGAAGKAGYPCTGVGPGPAAAAAAKAAAKTGA  242
Db  272  KAPKLPGGYGLPYTTNGKLPI-----GVAGAGSKASYFTGTGVGSQAAAAAA--KAAKYGA  324

Qy  243  GRAGVLPGVGGSAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAAAKKYGAAAGLVPGGPG  302
Db  325  GGAGVLPGVGGGIPGAGAIPIGGIAGAGTPAAAAAAAAAAAAKKAAKYGAAGGLVPGGPG  384

Qy  303  FGPGVVGVPGAGVPGV-GVPG-AGIPVVPGAGIPGAAV---PGVVSPEAAAKAAAKNAAKY  357
Db  385  -----VRLPGAGIPGVGGIPGVGGIPGPGGGVGGPGCAVSPAAAKAAAAKAAKY         439

Qy  358  GARPGVGVGGIPIYGVGAGGFPGPGVGVGGIPYVAGVPSVGVPGVGGVPGVGISPEAQA   417
Db  440  GARGGV---GIPTYGVAGGGFPGYGVAGA-----------GLGGA------SP-AAA     476

Qy  418  AAAAKAAKY-----------------------GVGTPAAAAKAXAAK-A             442
Db  477  AAAAKAAKYGAGGAGALGAGGLVPGGVPGAVPGGALPGAVPAVPGGGVPGAGTPAAAAAAAKAAKA  536

Qy  443  AQPGLVPGVGVAPGVGVAPGVGVAP---GVGLAPGVGVAPGVGVAPGVGVAP-GIGPGGV   498
Db  537  AKAGLGPGVGGVPG--GV--GVGGIPGGVGGVPG--GVGPG-GVT-GIGAGPGGLGGAGS   590

Qy  499  AAAKSAAKYAAKAQLRAAAGLGAGIPCLGVGVGPGVGACGVPGLGVGVGPGVPGGFGAGA  558
Db  591  PAAAKSAKAAKAAAQYRAAAGLGAGVPGFGAGAGVPGFGAGAGVPGFGAGAGVPGFGAGA  650

Qy  559  DEGVRRSLSPELREGGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAA----VPG       608
Db  651  --GV-----PGFGAG-----------AVPGGLAASKAAKYGAAGGLGAGAGGLGGPG    688

Qy  609  VLGGLGALGGVLPGGGVVGAG-PAAAAAKAAKAAQFGLVGAAGL----              654
Db  689  GLGGPGGLGGAGGLGGAGGLGAGGVSPAAAAAAAAKAAKAAQYGLGGAGGLGAGGLGAGG  748

Qy  655  ---------GGLGVGGLGVPGVGGLGGIPPAAAAKAYGAAGLGGVGGLGAGQPPLGGVA   705
Db  749  GLGAGGLGAGGLGAGGLGAGGLGAGGVSPAAAAKAAKYGAAGGLGGVL-GARPPFGGGVA  807

Qy  706  ARPGFGLSPIFP-------                                          731
Db  808  ARPGFGLSPIFPGGGAGGLGVGGKPPKPYGGALGALGGAGLQGGGCFGKSCGRKRK    860
```

RESULT 5
BART
elastin precursor - rat
N;Alternate names: tropoelastin
C;Species: Rattus norvegicus (Norway rat)
C;Date: 11-Jan-1991 #sequence_revision 16-Aug-1996 #text_change 22-Jun-1999
C;Accession: A36106; A30878; A36523; S02173; I54172; I68505
R;Pierce, R.A.; Deak, S.B.; Stolle, C.A.; Boyd, C.D.
A;Biochemistry 29, 9677-9683; 1990
A;Title: Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.
A;Reference number: A36106; MUID:91104868
A;Accession: A36106
A;Molecule type: mRNA
A;Residues: 1-864 <PIE>
A;Cross-references: GB:M60647; GB:J05292; NID:g207444; PIDN:AAA42269.1; PID:g207445
R;Deak, S.B.; Pierce, R.A.; Belsky, S.A.; Riley, D.J.; Boyd, C.D.
J. Biol. Chem. 263, 13504-13507, 1988
A;Title: Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.
A;Reference number: A30878; MUID:88330868
A;Accession: A30878
A;Molecule type: mRNA
A;Status: preliminary
A;Residues: 781-864 <DEA>
A;Cross-references: GB:J04035; NID:g207442; PIDN:AAA42268.1; PID:g207443
R;Franzblau, C.; Pratt, C.A.; Paris. B.; Colannino, N.M.; Offner, G.D.; Mogayzel Jr.,
J. Biol. Chem. 264, 15115-15119, 1989
A;Title: Role of tropoelastin fragmentation in elastogenesis in rat smooth muscle cel
A;Reference number: A36523; MUID:89339327
A;Accession: A35523
A;Molecule type: protein
A;Residues: 22-31 <FRA>
R;Rich, C.B.; Foster, J.A.
Arch. Biochem. Biophys. 268, 551-558, 1989
A;Title: Characterization of rat heart tropoelastin.
A;Reference number: S02173; MUID:89117149
A;Accession: S02173
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: '1P',369-545,548-764,770-864 <RIC>
A;Experimental source: heart
R;Pierce, R.A.; Alatavi, A.; Deak, S.B.; Boyd, C.D.
Genomics 12, 651-658, 1992
A;Title: Elements of the rat tropoelastin gene associated with alternative splicing.
A;Reference number: I54172; MUID:92241859
A;Accession: I54172
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 264-533 <RES>
A;Cross-references: GB:M86372; NID:g207455; PIDN:AAA42271.1; PID:g554527
A;Accession: I68505
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 558-864 <RE2>
A;Cross-references: GB:M86376; NID:g207459; PIDN:AAA42272.1; PID:g207462
C;Genetics:
A;Introns: 277/1; 292/1; 308/1; 339/1; 359/1; 419/1; 437/1; 467/1; 484/1; 601/1; 621/
A;Note: the list of introns may be incomplete
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-864/Product: elastin #status predicted <MAT>
F;854-859/Disulfide bonds: #status predicted

Query Match   61.1%;   Score 2311;   DB 1;   Length 864;
Best Local Similarity   58.5%;   Pred. NO. 9.3e-100;
Matches 531;   Conservative   26;   Mismatches   102;   Indels 248;   Gaps   34;

```
Qy    1  GGVPGAIPGGSVPGGVFYPGAGL-GALGGGALGPGGKPLKP--------VPGGLAGAGL   49
Db   30  GGVPGGLPGGVPGGVYYPGAGIGGLGGALGPGGKPKPKPGAGLLGFPGAGPGGLGGAGP   89

Qy   50  GAGLGAFPAVTFPGALVPGGVADAAAAYK-AAKAGAGLGGVPGV--------------   92
```

A:;Residues: 1-212,237-524,535-784 <BR2>
A:;Cross-references: GB:M15089; NID:g2128803; PIDN:AA49108.1; PIDN:AA449108
R:Baule, V.J.; Foster, J.A.
Biochem. Biophys. Res. Commun. 154, 1054-1060, 1988
A:;Title: Multiple chick tropoelastin mRNAs
A:;Reference number: A30795; MUID:88309083
A:;Accession: A30795
A:;Molecule type: mRNA
A:;Residues: 85-784 <RAD>
A:;Cross-references: GB:M21880; NID:g212741; PIDN:AAA49082.1; PID:g212742
R:Tokimitsu, I.; Tajima, S.; Nishikawa, T.; Tajima, M.; Fukasawa, T.
Arch. Biochem. Biophys. 256, 455-461, 1987
A:;Title: Sequence analysis of elastin cDNA from chick aorta and tissue-specific trans
A:;Reference number: A27264; MUID:87297534
A:;Accession: A27264
A:;Molecule type: mRNA
A:;Residues: 491-569,'G',571-604,'A',606-643,'A',645-687,'R',689-700,'R',702-784 <TOK>
C:Superfamily: elastin
C:Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F:1-24/Domain: signal sequence (fragment) #status predicted <SIG>
F:25-212,237-524,535-784/Product: elastin #status predicted <MAT>
F:773-779/Disulfide bonds: #status predicted

Query Match   44.8%; Score 1696.5; DB 2; Length 784;
Best Local Similarity 51.1%; Pred. No. 1.5e-71;
Matches 450; Conservative 38; Mismatches 123; Indels 269; Gaps 54;

Qy   1 GGVPGAIP-GGVPGGVFYPGAGLGALG---GGALGPGGKPLKPVPGGLAG----------- 46
             ||| |:|| |||| ||: | |: ||||| | ||||||
Db   25 GGVPGAIPGGGVPGGGFPPGAGVGGLGAGLGAGLGAGGKPLKPGVSGLGGLGPGLGLGPGA 84

Qy   47 -AGLGAGLGAFPAVTFPGALVPGGVTADAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVP 104
            |||||||||||| | ||: | |::|||||:|:||| ||:|| ||
Db   85 GVGGLGAGLGAFPGAFPGA------ASAALKAAAKAGAGLGGVGIGGLGGVGGVGVP 138

Qy 105 --------QPGAGV--KPGKVPGVGLPGVP-GGVLPGA--RPPGVGVLPGVPGVPTGAGVK 150
                     |:||| ||:|||||| : :|: :||||| :|||| || ||||:
Db 139 GGLGVPGVVQPGVGAGKPPKVPGAGIPGAFPGGVLPGGGIRPPGVGVLPGVPTGTGIK 198

Qy 151 PKAPGVGGAPAGIPGVPGGPGOPGVPLGYPGSYGLPGGYGLPYTGKLPTYGPGGV---- 209
             : || || |:| ||:|| || ||||| ||||| || :|| ||||||||||
Db 199 AKGPG-AGAFPAGIFGLGGSGGQDPGVPPGLGYIKAPLFGGTRLRPFVN-----GLGPGGIG 252

Qy 210 AGA-AGKAGIPTGIGVGPOAAAAAAKFGAGAGVLPGVGGAGVPGVPGAIPGIGG 268
            || || |:||||| ||||||||||||| ||| | |:|:|:||| ||| ||
Db 253 AGVLAGKAGYPTGTGVGAG----AAAKAAAKYG--AGVLPGAG--GIPGVGGVPGVGV 304

Qy 269 I -AGVGTPAAAAAAA-AAAKAAKYGAAAGLVPGGPGPGPGVSVPGVGGV-VPGAG 324
           : |||||:|||||||| ||||||||||||||||:|: ||||:||:|: ||||:||||
Db 305 VPGAGVGGPAAAAAAAAAAAAKAAAKAAKAAGAYGA--------GVLPGAGGPGV--VPGVVPG-- 352

Qy 325 IPVVPG-AGIPGAAVPGVVSPEAAAKAAAKAAKAAKYGAGARPGVCVGGIPTYGVGAGGPPGF-G 382
             || ||||| |:| || ||:|||||||||||||||||| : ||||||||| ||:|||| :|||| |
Db 353 -LVPGVGGIPGVA--GVGTPAGAAAAAAKAAKYGA---GVPGVGVPGVGIGGVPGVGVPG 404

Qy 383 V-GVGGIPGVAGVPGSVGPGVPGVGV-GISPEACAAAAAKAAKYGTP-AAAAKAA 439
            | || ||||||||||| ||||| |||||| :|| |||||||| ||||||||||
Db 405 VPGVGVPGVPGVPGVGVPGVPGVPGVVPGV--------GVGGPAAAAAAKAA 453

Qy 440 AKAAQPGL--VPGVGVAPGVVAPGVVAPGVCGLAPGVTAPGVCGLAPGVVAPGIGPGG 497
            |||| ||| |||||||:|| ||| |||: |: |||:|||:||||||||| ||| |:|
Db 454 AKAAAFGAGKRVPGVGP--GAVPGVGV-PGVGV-PGVGV-PGVGVPGG 505

Qy 498 VAAAKSAAKVAAKAQLRAAGGLGAGIPGLGV-GVGVPGLGV-GAGVPGLSVGAGVPGFG 555
            : |:|||||| |||:||||| ||||| ||||||||| |||||||||| ||| ||
Db 506 V-------GVPGVGVPGVGVPGVGVPGVGVPGVLVPGAG 541

Qy 556 AGGADEGVRRSLSPELREGDPSSSOHLPSTPSSPRVPGALAAAKAAKYGAA--VPGV---- 609
            :||||||||||||||||||||| ||||||||||||||||:||||||||||| ||||
Db 542 PAA--------------AAKAAAKAAKYGAGGLAPGVGGGLA 568

RESULT 6
A26601
elastin precursor - chicken (fragment)
N:Alternate names: tropoelastin
C:Species: Gallus gallus (chicken)
C:Date: 05-Oct-1988 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
R:Bressan, G.M.; Argos, P.; Stanley, K.K.
Biochemistry 26, 1497-1503, 1987
A:;Title: Repeating structure of chick tropoelastin revealed by complementary DNA cloning
A:;Reference number: A26601; MUID:87242320
A:;Accession: A26601
A:;Molecule type: mRNA

Db   90 GAGLSY-ASRPGGVLVPGGAGGAGAAAAYKAAAKAGAGLGGIGGVPGGVCGVCGPGAVGVG 147
             ||| | | | | ||| | ||
Qy   93 GGLGVSAGAVVPGAGV---   KPGKVPGVGLPGVPGGVTPGGVLP--GARP 134
             ||||| ||||:||||||||| |||| ||||:| ||||
Db 148 GVPGAVGGIGGIGGLGGLGVSTGAVVPGLGAGVGGAPGPGKVPGVGLPGVYPGGVLPGTGARP 207

Qy 135 PGVGVLPGVPGVPGAGVMPKAP-GVGGAPAGIPGVGGPGQPGVPLGYPIKAPKLPGGYGL 193
            ||||||||||||||||||||| || |:|||| |:||| || ||||| ||||||||||:||
Db 208 PGVGVLPGVPGVPTGVKAKVPGGGGGAFPGGIPGVGGPGGQPGVPLGYPIKAPKLPGGYGL 267

Qy 194 PYTTGKLPYGYGPGGVAGAAGKAGYPTGTGVGPGPAAAAAAKAAAKPGAAGVLPGVGG 253
             | |||| |:||| | |||||:|| |||||| :|||||||||||:|||||||||:||
Db 268 PYTNGKLPY----GVAGAGGKAGYPTGTGVGSG-AAVAAAAKAAKYAGAGGGVLPGVSG 321

Qy 254 AGVPSVPGAIPGIGGIAGVYGSTPAAAAAAAKAAKYGAAAGLVPGGPGPGGVPGVPGA 313
            |||||||||||||||||||| |||:||||||:|||||||||:|: | :||||||
Db 322 GGIPGGAGAIPGIGGSITGAGITGAGTPAAAAAKAAKYGAAGGLIVPGGPG------VRVPGA 376

Qy 314 GVPGVGVPG-------------AGIPVVPGAGIPGAAV---PGVVSPEAAAKAAAKAAKY 357
            ||||||| | | | |:   : |:| |||||||||||
Db 377 GIPGVGGIPGVGGIPGVGGICGPGIVCGGPGIVCGGPGGAVSPAAAAKAAAKAAKY 436

Qy 358 GAPGVGVGGIPTYGVGAGGPGTGVGV----------   385
             | | |||| |||| ||||
Db 437 GARGGV---GIPTYGVGAGGPGYGVGAGLGGASQAAAAAAAKAAKYGAGGAGIGG 493

Qy 386 -------GGIPGV-AGVP----------SVGGVPGVGGVPGVGISPEAQAAAAAKAAKYG 427
                    ||||| |||| :|| |||||||||| | :||||||||||
Db 494 LVPGAVPGALPGAVPGGALPGAVPGGVPGGVPGGVPGGVPGAGVAPGVGVAPGVGVVAPGV--TPAAAA 542

Qy 428 VGTPAAAAAKAAKAAQPGLVPGVGVAPGVGVAPGVGVAPGVGVAPGVGGAGVLPGV 487
            || |||||||||||| | |||||||||||||||||||||:|||||| | |||
Db 543 ---AAAAKAAKAAQCYLGPGVYGGVPG-GV--GVGGLPG-GVGPG-GVT-GIGTGPGT 592

Qy 488 GVAPG-IGPGGVAAAAAAKSAAKVRAKAQLRAAAGLGAGIPGLGVGVGVPGLGVGAGVPGLG 546
            ||| | || ||:||||| |:||| ||| |||||:||||||||||||||||||||||||||
Db 593 GLVPGSDLGGAGTPAAAK SAAAKAAKAQYRAAALGAGLGAGVPGGAGAG--GFG 650

Qy 547 YGGAGVPGFGAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAA- 605
            :|||||||||||| :|::||||||:|||||||||||||||
Db 651 AGAGVPGFGAGA--------   VPGSLAASKAAKYGAAG 679

Qy 606 ---VPGVLGGLGALGGV----------   GIPGGVGAGPRAAAAAKAAAKAAQFGLVGAA 652
               |||||||| ||:|| || |||||||||||||||||||||::|||:|
Db 680 GLGGPGGLGGLGPGLGGPGGPGGGPGGGAP- AAAAAAKAAAAKAAKAAAQGLGGAGG 738

Qy 653 GL----------   GGLGVGLG---VPGVGGLGGIPGAVGLGGVSPAAAAAAKYGAGLGG 691
             | |||| |||| ||:|||||| ::::||||||||||||||||||||||
Db 739 GLGGAGGLGAGGLGAGGLGAGGLGAGGVIPGAVGLGGVIPGAVGLGGVSPAAAAAKYGAGLGG 798

Qy 692 VLGGAGQFPLGGVAARPGPGLSPIFP----   VLGGAGLGK 724
            :||| ||||||| |:| | ::| || ||||||
Db 799 VL-GARPFGGGVAARPGPGLSPIYPGGGAGGLGVGGKPKPYGGALGALGVYQGGGCFGK 857

Qy 725 ACGRKRK 731
            :|||:||
Db 858 SCGRKRK 864

```
Qy  610  -LGGL----GALGGVG--IPG----GVVGAGPAAA--------AARARAAKRAAQF  646
              :||             ||    ||           ||||||||||||::
Db  569  PANVGGLAPGVGGLYP2VGGLVPGVGGLAPGVGAVPGVGGPAAAKAAAAKKAAKY  628

Qy  647  GLYVGAAGLGGL----------------GVG------GLGVPGVGGL--GGIPPAAA  678
                 :|:|:                 ||    ||   :||  | |  ||
Db  629  G----AGVGGVPGAVPGANVPGVPGVTPGVGGVPGVPSLVPGVGGVLPGVLPGAGIPPQVGV  684

Qy  679  AKAAK------YGAAGLGGVLG--GAGQPPLGGVAAR--------PGFGLSPIFPG-  718
              :|          |:||| :| | | :|| ||| :||       ||:|:|||:|:|
Db  685  QPGAXPPKFGVPGAGVPGVGGIPPGGLGVGGLGAGLGAGVPGFGVSPIFPGGV  744

Qy  719  ------GACLGKACGRKRK  731
               |||  | |||||
Db  745  GGGQLGFGGKGKPKIYGGAJGALGFRGGVGCAOGKYCGRKRK  784


RESULT  7
T29350
hypothetical protein F01G12.5a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
[ac]cession: T29350
R;Wu, X.; Le, T.T.
submitted to the EMBL Data library, April 1996
A;Description: The sequence of C. elegans cosmid F01G12.
A;Reference number: 220611
A;Accession: T29350
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-1758 <WUX>
A;Cross-references: EMBL:U53342; PIDN:AAA96215.1; GSPDB:GN00028; CESP:F01G12.5a
A;Experimental source: strain Bristol N2; clone F01G12
C;Genetics:
A;Gene: CESP:F01G12.5a
A;Map position: X
A;Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 736/3
C;Superfamily: collagen alpha 1(IV) chain

Query Match       17.8%;   Score 675.5;   DB 2;   Length 1758;
Best Local Similarity  30.1%;   Pred. No. 2.3e-24;
Matches  284;   Conservative  64;   Mismatches 323;   Indels 273;   Gaps  50;

Qy    2  GVPG-AIPGGVPGGVFYPG---AGLGALGGGALGPGGKPLKPVPGGLAGAGLG-AGLGA  55
            ||  |||||||||||  |||     || |||   |||:| |:| |||  ||||
Db  634  GLPGPSGPVGPPGAPGVPGEKGDAGLPGLSG---KPGQDGGLPGLPGNKGEAGVGQPGQPG  690

Qy   56  FPAVTFPGAL-----------------VPGGVADAAAAYKAAKRAGAGLGGVPGVGGLVSAG---AV  102
            |  ||| |                  :|                 |:||| | ||||
Db  691  FPGAKGDGGLPGLPGTPGLQGMPGPBPAPENQVNPAPPGQPGLPGTKGEGGYPGRGE  750

Qy  103  VPGPGAGVKPGKVPGVGGLPGV-----YPGGVLPGAR-TPGVGVLPGVPTGAG--VKPKAP  154
              | | |:| |  |  |||| |     ||| ||:|    ||| :| ||| ||||  ||||:|
Db  751  VGQPGFPGLPGMKGDSGLPGPGLPGHPG--VPGDKGPSGGVPGLPGIPGPKGDVGNPGLP  808

Qy  155  GVGGAPAGIPGVGPPGGP----------------QPGVP--------LGYPIKAPKLPGGYGL  193
            ||||     ||| |||                 :||||        |||:  ||||||
Db  809  GLNGQ-KGBPGVPGPGPGPCGPSPGPPGLKGDAGLPGLFGIPGLESGQRGFP-GAPGLKGGDGL  866

Qy  194  PVTTGKILPYGYGPG--------------GVRAGAGKAGIYPYGYGV-------------------  224
            | :  ||| |||:|||                  :|  ||||
Db  867  PGLSGQPGY---PGEKGDAGLPGVPGRGBGSPGFGQDGLPGVPGMKGEDGLPGLFGVTGL  923

Qy  225  ---------GFQAAAAAAKAAAKPGAGAGV----------LPGVGG---AGVPGV  259
                      ||             |||                |||  | ||||||
Db  924  KGDLGAPGQSGAPGLPGAPGYPGKKGNAGIPGVPGFKGDGGLPGLPGLMGPKGEPGVPGM  983

Qy  260  PG---------AIPGIGGIAGV--------------GTPAAAAAAAKAAKAYGAAAGLVP  298
              |          :|::|:||                  :||          ||:||| |
Db  984  PGTPGMKGNGGLPGLPGLPGRDLSGVPGMKGDRGFNGLPGEKGEAGPAARDGQKGDAG--LP  1041

Qy  299  GGSPGF-GP-GVVGVPGAGVPG----VGVPGAGIPGVVPG-AGIPGAAVPGVVSPEAAAKA  351
            ||:||| ||  |::| ||| ||     :| || :| | || |||||:|  |  |:  |
```

```
Db 1042  GQPGLRGPGQPSGLP--GVPGGYKGETGLPGYGQPGQPGBRGLPG-  1086

Qy  352  AKAAKYGARPGVGVGGIPTY-GV-GAGGPPG------------------  FGVGVGGIPGVA  392
              |:|:|      ::|| | :  |:  | | | |                  | :|||||| |
Db 1087  -KAGRQGAPGSPGQDGLPGFPGNKGBSGYPGQDGLPGRDGLPGVPGQKGDLGQSGQPGLS  1145

Qy  393  GVPGSVGGVPGV---------GGYPGV----GIS---------  PEAQAAAAKAAKTGV  428
            ||:||| ||| |          ||:||     |:          |  ||||||| |||
Db 1146  GAPGLDGQPGVPGIRGDKGQGLPGIPGDRGKDGIPGQKGBNGYPGQPGLPGLGGBRKGFA  1205

Qy  429  GTPAAAAKAA-AKAAQFGL--VPGVGVAPGVGVAPGVGVAPGVGLAGVGLAPGVGVAP  485
            ||||||||          |    ||||| | ||||||||:||| |:|
Db 1206  GTPGPFGSLKGSPGYPGQCDGLPGIPGLKGDSGFPGQFGQEGLPGLSGEKGWGGLPGMPGQP  1265

Qy  486  GVGVAPGIGPGGVAAAAKSAKVAAKAQLRAAAGLGA--GIPGLGVGVPGGLGVGAGVP  543
            |:|| || ||||||||:||:||||||:|    ||||   || | | | |||||||||:||
Db 1266  GOSIAGPVGPPG--------------------APGLQGKDGFPGLPGQKGBSGLSGLPGAP  1306

Qy  544  GLGVGAGVPGF-GAGADEGVR------RSLSPEL--REGDPS-------  SSQHLPSTP  585
            ||||||||| |  ||||:| |        |:|||   |:| ||        |||| |:|
Db 1307  GLKGESSMPGFPGAKGDLGANGIPGKRGEDGLPGVPGRDGQPGIPGLKGEVGAGLPGQP  1366

Qy  586  SSPRVPG---------ALAAAKAAKYGAAVPGVLGGLGALGGVGIPG-----GVVGA-----  628
            ||| ||            |:|||||||  :|| ||| || | |:||       | ||
Db 1367  GFPGIPGLKGBSGLPGFPGAKGBAGFPSTPGVPGYAGEKGDGGLPGLPGRDGLPGGADGSPV  1426

Qy  629  ---GPAAAAAKAAAKAAAQFGLVGAAGL--GGLGVGGL-GVPAGVGGLGGIPPAAAAAKAAK  683
                |                |:||||| :  ||||| ||| ||||:||||||:|||||||||:
Db 1427  GFPGPGSGPGNLYSPGEKGLP-GLPGAPGLRGEKGNPGLDGPPGNDGPGLP------GQ  1478

Qy  684  YGAAGLGGVLGGAGQPPLGGVAARPGFGLSPIFPGGACLGKACG  727
            ||  |||  :||| | || || || ||  | ||| ||:| |:
Db 1479  RGNDGYPGAPGLSGBKGMGGLPGFPGLDGPGGQPGPGAPG  1522


RESULT  8
T29351
collagen alpha 2(IV) chain precursor let-2 - Caenorhabditis elegans
N;Alternate names: collagen alpha 2(IV) chain precursor clb-1
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T29351
R;Wu, X.; Le, T.T.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid F01G12.
A;Reference number: 220611
A;Accession: T29351
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-1759 <WUX>
A;Cross-references: EMBL:U53342; PIDN:AAA96215.1; GSPDB:GN00028; CESP:F01G12.5a
A;Experimental source: strain Bristol N2; clone F01G12
C;Genetics:
A;Gene: CESP:F01G12.5a
A;Map position: X
A;Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 265/3; 304/3; 359/3; 450/2; 73
C;Superfamily: collagen alpha 1(IV) chain

Query Match        17.88%;   Score 675.5;   DB 2;   Length 1759;
Best Local Similarity   30.1%;   Pred. No. 2.3e-24;
Matches 284;   Conservative  64;   Mismatches 323;   Indels 273;   Gaps  50;

Qy    2  GVPG-AIPGGVPGGVFYPG---AGLGALGGGALGPGGKPLKPVPGGLAGAGLG-AGLGA  55
            ||  |||||||||||  |||     || |||   |||:| |:| |||  ||||
Db  635  GLPGPSGPVGPPGAPGVPGEKGDAGLPGLSG---KPGQDGGLPGLPGNKGEAGVGQPGQPG  691

Qy   56  FPAVTFPGAL-----------------VPGGVADAAAAYKAAKRAGAGLGGVPGVGGLVSAG---AV  102
            |  ||| |                  :|                 |:||| | ||||
Db  692  FPGAKGDGGLPGLPGTPGLQGMPGPBPAPENQVNPAPPGQPGLPGTKGEGGYPGRPGB  751

Qy  103  VPCPGAGVKPGKVPGVGGLPGV-----YPGGVLPGAR-TPGVGVLPGVPTGAG--VKPKAP  154
```

A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17745.1; PID:g292
A:Experimental source: strain H37Rv
C:Genetics;
A:Gene: Rv3508
C:Superfamily: collagen alpha 1(IV) chain

```
Query Match             17.5%;  Score 663;  DB 2;  Length 1901;
Best Local Similarity   33.6%;  Pred. No. 9.3e-24;
Matches  270;  Conservative  22;  Mismatches  324;  Indels 188;  Gaps  40;

Qy    2 GVPGAIPGGV--PGGVTFYPGAGLGALG--GGALGPGGKPLKPV------PGGLAGAGLG   50
          ||  ||   |   ||  : ||  :| ||    | || | :|| |          | ||  :|
Db  119 GANNTTPGGAGRAGGLLPGNGGAGGAAPGQCACGPGG----PAGLWGNGGPGGAGGSGGG  174

Qy   51 AGLGAFPAVTPGALVPG-VADARAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQCPGAG  109
          | | ||  |   ||  |   |  :| :| ||  ||  |  |  |||| |||||
Db  175 TG------GAGGAGGWLFGVGGAGGVGGAGGGLIWGG--------------GAG  218

Qy  110 VKPGKVPGVLPGVYPGGVLPGARPP---GVGVLPGVPGAGVKPFAGVPGGRAFGIPGGV  166
          |||  |||  ||   ||    | |       |  | |  | : |||  |  ||||| | |
Db  219 ---GVGGAGGDTGGA--GGRAELLFGAGGAGGAGGAGGASTDGG----PGATGGTGGHGGV  264

Qy  167 GPPGGPQPGVPLGYPIKAPKLPGGYGLPFYTTGKLPYGVPGGVAAGAAGKAGTPTGTG--V  224
          || ||   |  ||      | |||    ||| |  | ||  |:| |    ||   |:
Db  265 GGDGG--------------WLAPGGAGGAG-----GQSGGAGGSAGSDGGALGGTGGTG  302

Qy  225 GPQAAAAAAAAKAAKFGAGAAGVL------PGVGGAGVPGVPGCAIPGIGGIAGVGCTPAA-  277
          || | || | | |  |      |        || ||    |  ||  |  |  : | ||
Db  303 GTGGAGGAAGGRGALLLGAGGQGGLGGAGGGVGGTGGAGGGDGVLGGVGGTGGKGGVGGVAGL  362

Qy  278 AAAAAAAAKRAAKYGAAAGLV-PGGPGFGPGVVGGVGPGAGVPGVGVPGAGIPVVPGAGIPGA  336
          |||   |  | | | ||  |        | |   ||  |  ||  | | | :| | ||
Db  363 GGAGGAAGQLFSAGGAGACVGVGTG---GCGGAGGAAGAGADAP-ASTGLTGGTVGRAGG  418

Qy  337 AVPGVVSPEAAAKAAAKAAKYGARPGVGVGIPTYGVGAGGFPGFGV--------GVG  386
          |||| |     | || | |  |      || | |    ||  |  |        |||
Db  419 A--GGVGGQGGNAIAG------GINGSGGAGGTGGQC-GAGGMGGSGADNASGIGADGGAG  470

Qy  387 GIPGVAGVPSVGVGVPCGVGGTVPGVGISPEAQALARAKAAKYGVGTP--AAAAAKAAAKAaQ  444
          ||||| |||  | |  |  | | ||| |    |  || |  | || |     | | || | ||
Db  471 GTGGNAGAGGAGGAGGTGGTGGV-----------VGAAGKAGIGTGGQGGAGGSAGTDATA  523

Qy  445 FGLVPGVGAPGVGVAPGVGVAPGVGLAPGVGVAPGIG---PGGVAAA  501
          |||  | | |  |   |  | | |    | || | |       ||  || |
Db  524 TGATGGTGFSGGAGGAGGGNTGVGGTNGSGGGQGGTGGAGGAGGPGLGVGAGVPGFGAGGGT  583

Qy  502 AKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGVPGLGVGAGVPGFGAGADEG  561
          |  | | |  | |  |||  |  ||  | | | |||  | | | | | |  |  ||
Db  584 GGTGGKGGA------GGAGGQGGSSGAG---GTNGSGAGGIGGQG-GAGGAG--GAGAD--  631

Qy  562 VRRSLSPELREGDPSSSGQHLPSTPSSPRVPGALAAAAKAAKYGAAVPGVLGGLG--ALGGV  619
          |  | | ||   | |      :|       ||| | | | | | | |   | | |    | |
Db  632 -----------NPTGIGGAGGTGGTGGAGGAGGAGGAGGEGGAGGNSGVGGTNGSCGAGGACKG  678

Qy  620 GIPGGVVGAGPA-AAAAARAAKAAQF------------------------GLVGAAGLG  655
          |||||  |||                                             |||  ||| |
Db  679 GGTGGVGGAGGAGAAAAAAGSAATGGAGAGGAAAKYGAAGLGGVLGGAQGPFLGGVA  738

Qy  656 GL-GVGGLGVP------------GVGGLGGIPPAAAAKAAKYGAAGLGGVLGGAGGQPFLGGVA  705
          || | ||  |             |   |  | |  |  | ||    |  |  |  ||    |
Db  739 GTGGAGGSGADNPGAGFGAGGAGGTGGAGAGGAGGAGGAT-GTGGTGGVVGATGSAGIGGAG  797

Qy  706 ARP---------GFGLSPIPPGG 719
          | |              ::||  | |
Db  798 GXGGDGGDGASGLGLGLSG-FDGG 820
```

RESULT 10
A36068
major ampullate fibroin protein - orb spider (Nephila clavipes) (fragment)
C:Species: Nephila clavipes
C:Date: 08-Mar-1991 #sequence_revision 13-Jan-1993 #text_change 09-Sep-1997

```
Db  752 VQDPGPGLPGWKGDSGLPGPGLPGHPG--VPGDKGEGGVPGLPGGIPGGPKGDVGNPGLP  809
Qy  155 GVGGAPAGIPGVGPFGGP----------LGYPIKAPKLPGGYGL  193
Db  810 GLNGQ-KGPGGVPGQGPGSPGFPGCLKGDAGLPGLPGTPGLEGQRGFP-GAPGLKGGDGL  867
Qy  194 PYTTGKLPYGVGPG---------GVAGAAGKAGTPTGTGV----------------  224
Db  868 PGLSGQPFGI---PGEKDAGLPGVPGREGSPFGFGQDGLPGVPGMGKGEDGLPGLPGVTGL  924
Qy  225 --------GPGAAAAAAKAAKFGAGAAGV---------LPGVGG----AGVPGV  259
Db  925 KGDLGAPGQGGAPLGGAPGIYPGMKGNAGIIPGVPGFKGDGGLPGLPGLNGPKGBPGVPGM  984
Qy  260 PG--------AIPGIGGIAGV-----------GTPAAAAAAAAAKAAKYGAAAGLVP  298
Db  985 PGTPGWKGNGGLPGLPGRDGLSGVPGQKKGDRGFWGLLPGEKGENGPAAARDGQKGDAG--LP 1042
Qy  299 GGSPGF-GP-GVVGVPGAGVPG---VGVPGAGIPVVPG-AGIPGRAAVPGVVSPEAAAKAA  351
Db 1043 GQPGLRGPGPGPSGLP--GVPGGLKGBYGLPGYGQPGQPGEKGLPG--IPG 1087
Qy  352 AKRAKYGARPGVGVGIPPY-GV-GAGGEPG---------------FGVGVGGGIPGVA  392
Db 1088 -KAGRGQAPGSSPGQDGLPGFPGMKGESGTPGQDGLPGRDGLPGVPGGQKGDLGQSGQPGLS 1146
Qy  393 GVPSVGGVPGV-----------GIS----------PEAQAAAAAKRAKYGV  428
Db 1147 GAPGLDGQPGVPGIRGDXGQQGLPGIPGRGMDGYPGQKGENGTPGQPGGLPGLGGGEKGPA 1206
Qy  429 GTPAAAAAKAA-AKAAQFGL-VPGVGVAPGVGVAPGVGVAPGVGVGLAEGVGVAPGVGVAP  485
Db 1207 GTPGFPGLKGSPGTPGQDGLPGIPGLKGDSGFPGQPGBGLPGLSGEKGMGGLPGZMPGGQP 1266
Qy  486 GVGVAPGIGPGVAAAAKSAKVAAKAQLRAAAGLA--GIPGLGVGVGVPGLGVGAGVP  543
Db 1267 GQSIAGPVGPPG------------APGLQGKDGFPGLPSGQKGESGLSGLPGAP 1307
Qy  544 GLGVGAGVPGF-GAGADEGVR---------RSL-SPEL--REGDPS-------SSGHLPSTP  585
Db 1308 GLKGESSGWPGFPGAKGDLGANGIPGCKRGEDGGLPGVPGRDGQPGIPGLKGEVGGAGLPGQP 1367
Qy  586 SSPXVPG-------ALAAAAKYGAAVPGVLGGLGALGGVGIPG-----GVVGN-------  628
Db 1368 GFPGIPGLKGEGGLPGFPGGAKGEAGSPG*PGVPGVTAGEKGDGGLPGPGRDGLPGARDGPV 1427
Qy  629 ---GPAAAAAAAKAAAKAAQFGLVGAAGL-GGLGVGGL-GVPGVGGLGLGGIPPAAAAAKAAK  683
Db 1428 GPPGPSGPQNLVEPGEKGLP-GLPGAPGLRGGKGMPGLDGPPGNDGPPGLP-----GQ 1479
Qy  684 YGRAGLGGVLGGAQGPPLGGGVAARPGFGLSPIPPGGACLGKACG 727
Db 1480 RGNDGYPGAPGLSGEKGMGGLPGFPGGLLDGQPGGPGAPGLPGAFG 1523
```

RESULT 9
F70806
hypothetical glycine-rich protein Rv3508 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70806
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, J.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: F70806
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1901 <COL>

C;Accession: A36068
R;Xu, M.; Lewis, R.V.
Proc. Natl. Acad. Sci. U.S.A. 87, 7120-7124, 1990
A;Title: Structure of a protein superfiber: spider dragline silk.
A;Reference number: A36068; MUID:90384959
A;Accession: A36068
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-718 <XUA>
A;Cross-references: GB:M37137; NID:g159711; PID:g159712
A;Note: the authors translated the codon GGT for residue 292 as Gln, GTA for residue 374

Query Match   17.5%;   Score 661.5;   DB 2;   Length 718;
Best Local Similarity   35.2%;   Pred. No. 5.1e-24;
Matches   263; Conservative   25; Mismatches 320; Indels 139; Gaps 30;

Qy   2 GVPGAIPGGVPGGVFYPGAGLGALGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVTF 61
            | ||| || | || ||| ||||||||
Db   4 GAAAAAAGGAGQG------GYGGLGPGGAGQGGY-----GYGGLGPGGAGQGGAGAAAAAA 51

Qy   62 PGA-----------LVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPGPGAG 109
             ||| ||| | | | || | | ||||||| ||| | |||||
Db   52 GGAGQGGYGGLGSQGAGRGGQGAGAA--AAAGGAGQGGYGGLGSQGAGRGGLGQG-GAG 108

Qy   110 VKPGKVPGVGLPGVYPGGVLPGARPPGVGYLPGVPTGAGVKPKAPGVGGAFAGIPGVGPF 169
             | | | | | |||| || | || ||| |||
Db   109 AAAAAAGGAGQGGYGGLGNQGAGRGGQGAGAAAAAGGAG-----QGGYGGLGSQGAGRGGL 164

Qy   170 GGPGPGVPLGYTPIKAPKLPGGYGLPYTTGKLPGGYGPGVGAGAGKAGYPTGTGVGPQAA 229
             | | || | | | || |||||| | || | || | ||
Db   165 GGQGAGAAAAAGGAGQ---GGYG------GLGSQGAGAQGGY-GGLGSQGAGRGG-GLGQGA 215

Qy   230 AAAAKAAAKFGAGRAAGVLPGVGGAGVPGVPGAIPGIGGIIAGVGTPAA------AAAAA 283
             || | |||||| ||| ||| || ||| | | | ||||| ||||
Db   216 GAAAAAAGGAGQGGYGGLGSQGAGRGGLGQGAGAAAAAGGAGQGGYGGLGSQGAGRGGGAAAA 275

Qy   284 AAKAAKYGRAAGLVPGGPGP--GVVGVPGAGVPGVGVPGAGIPVVPGAGIPGAAVPG 340
             ||| | | || |||||| ||| ||| ||| | |||| | | || |||| |
Db   276 AAGGAGQGGYGGL--GGQGAGAGQGGYGGLGGQGAGAGAAAAAGGAGQGGYGGLGSQG 333

Qy   341 VVSPEAAAKAAAKAAKYGARPGVGYGVGAGGPGFGVGVGGIPGVAGVYGGIVGGV 400
             | | || | | | | || | ||| || || | || || | || | |
Db   334 AGQGAGAAAAAAGGAGQG---GYGGLGSQGAGRGGLGQGAGAGAVAAAAGG----AGG 384

Qy   401 PGVGGVPGVGISPEAQAAAAAAKAKYGY--------------GTPAAAAAAAKAAK 441
             ||| | | | | ||| |||| | || | ||||| |
Db   385 GGYGGLGSQGAGRGGQGAGAAAAAGGAGQGGYGGLGNQGAGRQRGYGGLGNQGAGRGGLGQ 444

Qy   442 AAQPGLVPGVGGVGVAPGVGVGVAPGVGLAPGVGVGVAPGVGVAPGVGVAP-GSPGGVAAA 501
             | ||| || |||| | || | ||| | | |||||| ||| || | ||
Db   445 GAGQGGYGGLG--NQGAGRG------GQGAGAAAAGGA-GQGGYGGLG-SGGAGRGGQGAG 494

Qy   502 AKSAAKYAAKAQLRAAAGLGAGIPGLGVGVGVGYGAGVPGLGVGVAGGPVPGAGAGADEG 561
             || | | || || |||| | || ||||| | || ||| | ||| || | ||
Db   495 AAAAAAVGAGQG--GIRGQGAGQGGYG-GLGSQGSGSRG-GLGSQGAGA1AAAAAGGAGQGG 550

Qy   562 VRKSLSPELREGDPSSSQHLPSTPSSPRVPGALRAAAAKYGAAVPGV-LGGLGALGGGVG 620
             | || | | | || || ||| | |||| | | |
Db   551 L----------GGQGAGQ------------AAAGGVRGQGGYGGLGSQG 582

Qy   621 IPGGYVGAGPAAAAAAKAAAKAAQFGLVGAAGLGGLLGVGGLGVPGVGLGGIPRAAAAK 680
             ||| | || | | || | | || | | || ||||| | || ||
Db   583 AGRGGQGAGAAAAAA----------GGAGQGGYGGLGGQGVGVGRGGLGGQG-----AGAAAA 627

Qy   681 AAKYGAAGLGGVLGAGQPGLGGVAAR 707
             || | | ||| | ||| || | |
Db   628 AAGGAGQGGYGGLGSQGAGRGGLGGQG 652

S16366
collagen alpha 2(IV) chain precursor - pig roundworm

C;Species: Ascaris suum (pig roundworm)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C;Accession: S16366
R;Pettitt, J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A;Title: The complete primary structure of a nematode alpha-2(IV) collagen and the pa
A;Reference number: S16366; MUID:91340768
A;Accession: S16366
A;Molecule type: mRNA
A;Residues: 1-1763 <CB1>
A;Cross-references: GB:M67507; NID:g159648; PIDN:AAA18014.1; PID:g159649
C;Genetics:
A;Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disul
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1753/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F;27-42/Domain: non-collagenous NH1 #status predicted <NH1>
F;43-1529/Domain: collagenous #status predicted <COL>
F;197-199/Region: cell attachment (R-G-D) motif
F;1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F;1530-1638/Domain: repeat NC1 #status predicted <NC11>
F;1639-1763/Domain: repeat NC1 #status predicted <NC12>
F;31,34,39,41,536,539/Disulfide bonds: interchain #status predicted
F;126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match   17.0%;   Score 644;   DB 2;   Length 1763;
Best Local Similarity   31.2%;   Pred. No. 6.5e-23;
Matches   277; Conservative   58; Mismatches 291; Indels 262; Gaps 54;

Qy   2 GVPGAIPG--GVPGGVFYPGA----------------------------------GLGALGG-------- 27
            | ||| | || ||| ||| ||||||
Db   427 GIPGA-PGVQGPPGGLAGPPGAKGSPPGPRGVDGQSIPGLPGKDGRPGLDGLPGRKGEMGLP 485

Qy   28 GALGPGGKPLKPVPG------------------GLGAGLGAGLGAFPAVIFPGALVPGGVADAAAAYKA 79
            || | || | || | ||| | ||| || | | | | |
Db   486 GVRGPPGDSLNGLPGPPGPRGPPGPGPKGYDGRDGAPGLPGIPGPKGDRGGTCAFCAHGAKG 545

Qy   80 AKAGAGLGGVPGVGG----------LGVSAGAVVP----QPGAGVPKPGKVPGV-----GLPGV- 123
            | | ||| | ||| | | || || || | || | | || || ||
Db   546 EKGDAGYAGLPGPQGDRG-LPGIP-SATGAPGDDGLPGAPGRPGP-PGPPGQDGDGLPGLPGQGK 604

Qy   124 ----------------YPGGV----LPGAR---------FPGVGVLPGVPTGAGVK------------- 150
                             | ||| |||| ||||| ||| | ||
Db   605 GPPTQLQILRPGPPGCYPGQKGETGFPPGPRGQOEGLPGKPGIVGARGLPGPPGPKGEPGLLGL 664

Qy   151 PKAPSVGGAFAGIPGV-GPFIGPGPQPGVPLGYP-IKAPK-LPGGYGLPYTTGKL--------- 200
            | | | || | || || | ||| || | | ||| | |||| | | ||
Db   665 PEKFGKDG-IPGLPGLKGSPGYGQPGMP-GLPGMKGSDAGLPGLPGLPGGPGVPMGPPVPES 722

Qy   201 ---------PYGYGPGVAGAAGKAGYPTGTGV-GPQAAAAAAAAAAKFKGAGAGVLPGV-G 252
                       || || | | | || | || || | ||| || | || | | |
Db   723 QLRPGPPGKDGLPGLPGPKGEAGFPGAPGLQGP-----------------------AGLPG-LPGKMG 766

Qy   253 GAGVPGVPG--AIPGIGGIAGV-GTPAAAAAAAAAAAAAAKYGAAAGL-------------- 296
            | | |||| | || | | | || | || | | |||| | | |
Db   767 NPGLPGAPGLAGLPGIPGEKGLPGLHGAKGENG-YPGQGPLPGKGEPGPSTG 825

Qy   297 VPGGPGPG---------PGVVGVPG-------AGVPG----VGVPG-AGIPVVPGA------G 332
            | ||||| | | | | || ||| | ||| | || | | |
Db   826 PPGPGPGPLGKGKDGIPGAPGLPGLESQGQGLPGVPGQGKGEIGLPGLAGAPGPGAKGEPG 885

Qy   333 IPGAAVPGVVSPEAAAKAAAAAKAKYGARPGVVGVGQIPTGVGAGGPPG--PGVGVGGIPG 390
            ||| | || | || | | | || | | | | || || | | |||||||
Db   886 LPG--LPGKGPQ---------------------GPPGQP--------GAPGFPGQKGDEGLPGLPG 921

Qy   391 VAGVPSVGGVGVGGPGVGVGISPEAQAAAAAAKAAKYGVGTPAAAAAKAAAAKRAAQPGLVPG 450
            | | | | | || | | | | | | ||| || | | || |||
Db   922 VSGMKGDTGLPGVPGLAG----PPGQPGPGQ--KGQPGFGSVAGAKGEA-------GI-PG 969

Qy   451 VGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGIGPG----------GVAAAAK 503
            | | |||| | || | |||| | || | |||| ||

Db 970 LPGAPGQKGEQLAGLPGI---PGMKGAPGIPGAPGQDGLPGL-PGVKGDRGFNGLPGEK 1025

Qy 504 SRAKVAAKAQLRAAAGLAGAIPGLGVGVPGLGVGAGVPGLGVGAGVPGFGFAGADEGVR 563

Db 1026 GBPGPAARDGEKGEPGL-PGQPGLRGPGPPGL---PGL-PGLKGPGDEGQ2PGYGA------ 1074

Qy 564 RSLSPELRGDPGSSSQHLPSPPSSPRVPEGALAAKAAKYGAAVPGVLGGALGGVGIPG 623

Db 1075 ----PGL-----MGEKGLPGLPGKPGRPGA-PGPKGLDGAFGFPGLKGZAGLPGAPGLPG 1124

Qy 624 --GYVGAGPAAAAKAAKAAQPGLVGAAGL-GGLGVGGL------------GVPGVGGL 670

Db 1125 QDGLPG------LPGQKGESGFPGQPGLVGPPGLPGKMGAPGIRGSKDAGLPGLPGERGL 1179

Qy 671 GGIPPAAAAAKAAKYGAAGLGGVLGGAGQPPLGGVAARPGFGLSPIPPG 718

Db 1180 DGL-PGQKGEAGFPGAPGLPGPVGPKGS---AGAPGFPGLKGEPGLPG 1223

RESULT 12
T31328
fibroin - Chinese oak silkmoth
C;Species: Antheraea pernyi (Chinese oak silkmoth)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T31328
R;Sezutsu, H.; Tamura, T.; Yukuhiro, K.
submitted to the EMBL Data library, August 1998
A;Description: Characterization of the full length fibroin gene of a wild silkworm, Anth
A;Reference number: Z20995
A;Accession: T31328
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-2639 <SE2>
A;Cross-references: EMBL:AF083334; NID:g3450882; PIDN:AF083334; PID:g3450883; PIDN:AAC32606.1
C;Genetics:
A;Introns: 14/3

Query Match   16.8%;   Score 635;   DB 2;   Length 2639;
Best Local Similarity 29.8%;   Pred. No. 2.3e-22;
Matches 258;   Conservative 373;   Mismatches 373;   Indels 198;   Gaps 25;

Qy 5 GAIPGGVPGGVFY-----------------------------PGAGLGALGGGALGPGGKPLKPVPGGL 44

Db 1056 GSGAGGVGGGYGWDGGGYGGYGSDSAAAAAAAAAAAAAAGSGAGGRGDGGYGSGSSAAAAAAA 1115

Qy 45 AGA-----------------GLGAGLGAFPAVTFPGALVPGG------------------VADAAAYKAAK 81

Db 1116 AAARRAGHDRAAGSAAAAAAAAAAAAGSGAGGSGGGYGWGDGGYGSDSAAAAAAAAAA 1175

Qy 82 AGAGLAGGVPGVGGLGVSAGNVPQPGAGVKPGKVPGVGLPGVGVLPGARFPGGVTLP 141

Db 1176 AASGAGGSGGGYGYGSDSAAAAAAAAAGSCAG-GAGGGYGWG------DGGYGSDS 1228

Qy 142 GVPTGSAGVKPAPGVGGARFAGIPGVGPGGPQPGVPGLGYP1KAPKLPGGYGLPYTTGKLP 201

Db 1229 AAAAAAAAAAAAGSGAGGSGGGYGYGSDSAAAAA---AAAAAAGSGAGGAGG--G 1283

Qy 202 YGVGPGGV--------------------------AGAAGKGTVPTGTGVGFGPQAAAAAAA-------- 234

Db 1284 YGWGDGGYGSDSAAAAAAAAAAAAGSGAGGRGDGGYGSGSSAAAAAAAAAARRAGHD 1343

Qy 235 ----KAAAKFGAGRAAGVLPGVGGAG------------------VPGVPGAIPGIG 267

Db 1344 RAAGSAAAAAAAAAAAASGAGGSGGGYGGGYGWGDGGYGSDSAAAAAAAAAAAASGAGSG 1403

Qy 268 GIAGVGPGPAAAAAAAAKAAKYGAAAGLVPGGPGPGPGVVSGVAGVPGVPGAGIPV 327

Db 1404 GYGGYGSDSAAAAAAAAAAAA-GSGAGGVGGGYGWGDG-----GYGSDSAAAAAAAAAA 1457

Qy 328 VPGAGIPGAAVPGVVSPERAAKAAAAKAAKYGARPGVVSGVGIPTVGVGAGFPGFG--VGV 385

Db 1458 AAASGAGGSGGGYGYGSDSAAAAAAAAAAAAAAASGAG------------GAGGYGGYGGYGSDSRA 1508

Qy 386 GGIPGVAGVPSVGGVPGVYGGVPGVG-----------ISPEAAAAAAKRA----------KYGVG 429

Db 1509 AAAAAAAAAAAGSGAGGAGGDGGYGWGDGGYGGYGSYSAAAAAAAAAAAGSGAGGRGDGGYGSG 1568

Qy 430 TPAAAAAKAAKARQFGLVPGVGVAPGVGVAPGVGLA-PGVGVARFGVGVAPGVGV 489

Db 1569 SSAAAAAAAAAAAR------------------RAGHDRAAGSAAAAAAAAAAAAAAAASGAG-------------GAGG 1614

Qy 490 APGIGPGGVAAAAKSAAKVAAKAQLRAAAGLAGAIPGLGVGVGVPGLG----------------- 537

Db 1615 GYGWGDGGYGSDSAAAAAAAAAAA--AAAGSGAGGAGGGYGWGDGGYGSDSAAAAAAAAAAA 1671

Qy 538 -VGAGVPGLGVGAGVPGFGAGADEGVRRSLSPELRSG2PSSSQHLPSTPSSPRVPGALA 595

Db 1672 AAAGSGAGGRGGGYGWGDGGYGSDSAAAAAAAAAAAASGGGSYGWGDGGYGSDSAAAAAAAAARAAAA 1731

Qy 596 AAKAA--------------KYGAAVPGVLGGLGALGGVGIPPGVVGAGPA-----AAAAARAKAK 642

Db 1732 AAAAAARRAGHDRAAGSAAAAAAAAAAAAAAAASGAGGGSYGWGDGGYGSDSAAAAAAAARAA 1791

Qy 643 AAQPGLVGAAGLGGLGVGGLGVPGVGLGGIPPAAAAKAAKYGAAGLGGVLGGAGQFPLG 702

Db 1792 AAASGAGGSGGGYGY6-----------GYGGYGSDSAAAAAAAAAAAAAAGSGAGGVG---L--G 1840

Qy 703 GVA-ARPGFGLSPIFPGGACLGKACG 727

Db 1841 GYGWGDGGYGSDSAAAAAAAAAAAAAG 1866

RESULT 13
D70807
hypothetical glycine-rich protein Rv3514 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: D70807
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.;
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: D70807
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1489 <COL>
A;Cross-references: GB:AL022022; GB:AL123456; PIDN:CMA17751.1; PIDN:CMA17751.1, PID:g292
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3514
C;Superfamily: collagen alpha 1(IV) chain

Query Match   16.5%;   Score 626;   DB 2;   Length 1489;
Best Local Similarity 32.7%;   Pred. No. 3.8e-22;
Matches 260;   Conservative 26;   Mismatches 347;   Indels 162;   Gaps 35;

Qy 2 GVPGAIPGGV--PGGVFYPGAGLGALG--GGALGPGGKPLKPV-------------PGGLAGAGLG 50

Db 119 GANATTPGGRAGGAGGLLFGNGGAGGAGAPGQAGGPGG----PAGLWGNGGPGGAGGSGGG 174

Qy 51 AGLGAFPAVTFPGALVFGG-VADAAAAYKAKAGAGGLGGVPGVGLGVSAGAVVPQPGAG 109

Db 175 TG---------GAGGAGGNLFGVGGAGGVGAGGSTGGAGGPGGLIWGG------GAG 218

Qy 110 VKPGKVPGVGLPGVYPGGVLPGARFP---GVGVLPGVPTGAGVKPKAPGVGGAFRAGIPGV 166

Db 219 ------GVGGAGGGTGGA--GGRAELLFGAGGAGGAGNTDGG----PGATGGTGGHGGV 264

Qy 167 GPFGGPGFGVPGLGFIPIKAPKLPGGYGLPYTTGKLPYGYGPGGVAGRAGRAGTPTGTG---V 224

```
Db  265 GDDGG------------------GQGGAGGAGSDDGALGGGIGGIG 302
Qy      ------WLAPGGAGGAG-----GQGGAGGAGSDGGALGGGTGGTG 70

Db  225 GPQRAAAAAAKAAKFGAGAAGVL------PCVGGAGVPGTVGGAIPGIGGIAGVGTPAAA 278
Qy  53  LGAPPAVTFPGALVPGGTVADAAAAYKAAXAG------AGLGGVPGVGGLGVSAGAV 102

Db  303 GTVGGAGGAGGALLGAGGQGGLGGAGGQCGTPGAGGAGCDCVLGGVGGVGTGGKGGVGGVAGL 362
Qy  71  SGA--------AGGAGGAGGAGGLFASSGGSGGGFGFASTGGAGAGLFASGG-- 120

Db  279 AAAAAAKAAKYGAAAGLVPGGPGPGFGVVGVFGGAGVEFGVGVPGGAGIPVVPGAGIPGAAV 338
Qy  103 VPQPGAGVPGKVPGVGLPGVPGGVLFGGARPGVGVLPGGVPTIGAGYKPKAPGVGGA--F 160

Db  363 GGA-------GGAAGQLFSASG-AAGNAGVGGAGGAGGQG-GDCGVAG-----GAG-ADADQ 405
Qy  121 VGGTGGGAGSGGTCGVVGGTG-GAGGLFASGGAGGAG--------GSGCTGGGAGGTGGAAGL 172

Db  339 PGVVSPEAAAKAAAAKAKYGARPGV------GVGGIPPYGVGAGGFPGPGV------G 384
Qy  161 AGIPGVGPFGGP---------QPGVPLGYPIKAPKLPGGVGLPYTTGKLPYTGYGPGGVAGA 212

Db  406 PGATCGTGFAGGGAGGGAGGSSGAGGTNGSSGAGAGGGQG-GAGGAGGAGADNPTGIGGTGG 464
Qy  173 FGAGGAGGLGGQGNHTGGHGGAGGSAGLLALGDGGAGGAGGAGGAATT------GTGGAGGA 225

Db  385 VGGIPGVAGVPSVGGVPGSVGGVPGVGVGISPEAGAAAAAAKAAKYGVGTPAAAAAAAAKAAQ 444
Qy  213 AGKAGYPTGT-GVGPGAAAA------AAAKAAKFGAGAAGVLPG------------- 250

Db  465 DGGTGGAAGGAGAGGAGTGGTGGN-IGTTGNAGVGGAGGQGGGDG----GAGGAGGADADQ 519
Qy  226 GGKAGLLFGSSGAGGSGGAAGTEGUTGNSGGAGGAGGAGGKAGLLFGSSGAGGSGGAGGFANG 285

Db  445 FGLVPGVGVAPGVGVAPGVGVAPGVGGAPGVGVAPGIG---PGGVAAA 501
Qy  251 -VGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAKAAKYGAAAGLVPGGPGFGPGPGVTG 309

Db  520 PGATGGTGFAGGAGGAGGAGGSSGAGGTNGSSGAGGTSGCGGAGGAGGAGGADNPTGIGGT 579
Qy  286 STGGAGGAGGGAGLLGNGNGGSGGTSVATCGAGNGGAGGAGGLLIGNGNGGSGGGAGGGGGM 345

Db  502 AKSAKVAAKAQLRAAAGLGAGIPGLGVGVGVPG-LGVGAGVPGLGVGAGVPGTGAGADE 560
Qy  310 -VPGAGVPGVGVPGAGIPVVPGAGIPGAAVAPVGVVSPEAAAKAAAAKAKYGARPGVGVGGI 368

Db  580 GGDGGTGGAAG-AGGAGAAGTGGTGGMIGTTGNAGVG-GAGGQGGDGGAGGAGGADADQ 636
Qy  346 DAPG----GTGVGGIGGILL--GLGGANAPASTNPLHTACQQGALAVNA--PIQAVTGR 396

Db  561 ------GVRRSLSPELREEDPSSSQHLPSTPSSPRVGGPGALAAAKAAKYGAAVPGVLGGL 613
Qy  369 PTYGVGAGGEFPGKNGVGVGGIPGVAGVPSVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGV 428

Db  637 PGATGGTGFAGGAGGGAGGSSGAGGTNSSGSSGAGGTGRQSTGGAGGAGADNPTGIGGT 696
Qy  397 PLIGNGANGAPG----SGAPGGHG---GWLFGGGTGGSGGVSGGGAGGDGGAGGILLFGA 447

Db  614 GALGGVGIPGGVVGAGPAAA----AAAKAAAAKAAQFGLVGAAGLGGL---657
Qy  429 GTPAAAAAAAKAAAAQPGLVPGVGVAPGVGVGVGVGVAPGVGLAPGVGVAPGVG--VAPG 486

Db  697 GGDGGTGGAGGAGGAGGAAGTGGTGGMIGTTGNAGVGGAGGSSGAGGTDCQ 756
Qy  448 GGAGGAG------GAVTGTGGSSGGAGGAGLLFGAGGAGGAGGSSGIGGFRAGG 497

Db  658 ----GVGG-------LGYPGVGLGGIPPAAAAKAA-KYGRAGLGGVLGGRAGQFFPLGGVTA 705
Qy  487 VGVAPGIGPGGV-------AAAAKSRAKVAAKAQLRAAAGLAGIPGLGVGVGVPGL-GV 538

Db  757 GGAGGAGGADNPTGIGGTGGDGDGGTGGAAGGAGGAGAAGTGGTGGMIGTTGNAGVGGAG 816
Qy  498 AG------GPGGAGGLFNGGGAGGAGGSGVSGGAGGGAG-GAG--GLFAGGGAGGAGGS 549

Db  706 ARPGFGLSPIFPGGA 720
Qy  539 GAGVPGLGVGAGVPG-FGAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGALAAA 597

Db  817 GQGGDGGA----GGA 827
Qy  550 GNNYGGAGGAGGVGGLFGAGGAGG------------SGGGGS----------VAGDSGAG 587

Qy  598 KAAKYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAAKAAKAAQPGLVGAAGLGGL 657
Db  588 GNA--GLLAPGLAGGAGGGGGGQGFDTGGAG-GPGGDAGL------------LVSGGVGGA 633

Qy  658 GVGGL--GVPGV--------GGLGGIPPAAAAKAAKTGAAGLGGVLGGAGQFPLGGV 704
Db  634 GGFGLITTGGPGGAGGADGGDAGLLFGSGGAGG--AGGSGRTDLGGAGGAGGKAGLITGNGNGGA 691

Qy  705 AARPGFGLSPIFPGGACLGKACG 727
Db  692 GGAGGNGGDGGDGPGGAATGLGNG 714
```

RESULT 14
A70812
hypothetical glycine-rich protein Rv0833 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70812
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
andream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, J.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID: 98295987
A;Accession: A70812
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-749 <COL>
A;Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAA17639.1; PID:g3261689
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0833
C;Superfamily: elastin

Query Match   16.1%; Score 611; DB 2; Length 749;
Best Local Similarity   32.4%; Pred. No. 1.1e-21;
Matches 250; Conservative 30; Mismatches 337; Indels 176; Gaps 35;

```
Qy    2 GVPGALPG-GVPGGVFYPGAGLGALG------GGALGPPGGAFGPGKLPKPVPGGLAGAGLGAG 52
          |  ||||| || |     |        ||| |      || ||| | |
```

RESULT 15
CGHU4B
collagen alpha 1(IV) chain precursor - human
N;Alternate names: procollagen alpha 1(IV) chain
C;Species: Homo sapiens (man)
C;Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 07-Dec-1999
C;Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863;
R;Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 264, 13565-13571, 1989
A;Title: Structural organization of the gene for the alpha-1 chain of human type IV c
A;Reference number: S16876; MUID:89340433
A;Accession: S16876
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1669 <S01>
A;Cross-references: EMBL:J04217; GB:J05039; NID:g180800; PIDN:AAA53998.1; PID:g180803
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988
R;Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 263, 17217-17220, 1988
A;Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are
A;Reference number: A92690; MUID:89034231
A;Accession: A32117
A;Molecule type: DNA
A;Residues: 1-28 <SOI2>
A;Cross-references: EMBL.J04217; NID:g180759; PID:g553233
R;Poeschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A;Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c
A;Reference number: S02738; MUID:89030632
A;Accession: S02738
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-6,'L'/,8-28 <POE>
R;Brazel, D.; Oberbaeumer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.;
·.,. J. Biochem. 168, 529-536, 1987
~itle: Completion of the amino acid sequence of the alpha1 chain of human basement mem
·reference number: S00048; MUID:88029471
A;Accession: S00048
A;Molecule type: mRNA
A;Residues: 1-318, 'A',320-944 <BRA1>
A;Cross-references: EMBL.X05551; NID:g30066; PIDN:CAA29075.1; PID:g30067
A;Accession: S25826
A;Molecule type: Protein
A;Residues: 271-318,'A',320-554 <BRA2>
R;Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.
Eur. J. Biochem. 152, 213-219, 1985
A;Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7S
A;Reference number: A23115; MUID:86004708
A;Accession: A23115
A;Molecule type: protein
A;Residues: 28-236,'KE'/,239-240,'K',242-243 <GLA>
A;Experimental source: placenta
A;Note: the amino end of the mature form is blocked
R;Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.
FEBS Lett. 225, 188-194, 1987
A;Title: Complete primary structure of the alpha(1)-chain of human basement membrane (ty
A;Reference number: S00207; MUID:88083584
A;Accession: S00207
A;Molecule type: mRNA
A;Residues: 244-530 <SOI3>
A;Cross-references: EMBL.Y00706; NID:g29548; PIDN:CAA68698.1; PID:g29549
R;Eble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
EMBO J. 12, 4795-4802, 1993
~le: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen
~ference number: S39614; MUID:94038963
A;Accession: S39614
A;Molecule type: protein
R;Babel, W.; Glanville, R.W.
Eur. J. Biochem. 143, 545-556, 1984
A;Title: Structure of human basement-membrane (type IV) collagen. Complete amino-acid se
A;Reference number: A02863; MUID:85003629
A;Accession: A02863
A;Molecule type: Protein
A;Residues: 534-718,'D',720-836,'Y',838-841,'P',843-903,'K',915-997,'K',999-
A;Experimental source: placenta
R;Glanville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
A;Title: Pepsin fragments of human placental basement-membrane collagens showing interru
A;Reference number: S16908; MUID:82005835
A;Accession: S16908
A;Molecule type: protein
A;Residues: 534-537,'G',539,'G',541-542,'X',544-553,1389-1405,'XX',1408-1409,'X',1411-14
R;MacWright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
Biochemistry 22, 4940-4948, 1983
A;Title: Isolation and characterization of pepsin-solubilized human basement membrane (t
A;Reference number: S16910; MUID:84053346
A;Accession: S16910
A;Molecule type: protein
A;Residues: 534-537,'G',539,'G',541-542,'G',544-549,939-940,'M',942-944,'V',946,'X',948-

A;Experimental source: placenta
R;Pihlajaniemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.
J. Biol. Chem. 260, 7681-7687, 1985
A;Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollage
A;Reference number: SC14466; MUID:85207819
A;Accession: S01466
A;Molecule type: mRNA
A;Residues: 1256-1669 <PIH>
A;Cross-references: EMBL.M10940; NID:g180421; PIDN:AAA52006.1; PID:g180424
R;Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
A;Title: Restricted homology between human alpha-1 type IV and other procollagen chai
A;Reference number: S16879; MUID:85216555
A;Accession: S16879
A;Molecule type: mRNA
A;Residues: 1259-1669 <BRI>
A;Cross-references: EMBL.M11315; NID:g180817; PIDN:AAA52042.1; PID:g180818
R;Oberbaeumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Vos
Eur. J. Biochem. 147, 217-224, 1985
A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alph
A;Reference number: A02864; MUID:85127033
A;Accession: SI9091
A;Molecule type: protein
A;Residues: 1435-1661,'H',1463-1482,'X',1484-1491;1501-1514,'X',1516-1519;1534-1553,'
R;Siebold, B.; Deutzmann, R.; Kuehn, K.
Eur. J. Biochem. 176, 617-624, 1988
A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyt
A;Reference number: S02550; MUID:89005112
A;Contents: annotation; disulfide bonds
C;Genetics:
A;Gene: GDB:COL4A1
A;Map position: 13q34-13q34
A;Cross-references: GDB:119791; OXIM:120130
A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 2
/1; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 999/1; 1020/1; 1066/3; 1109/1; 1136/1;
C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha
ociations among trimer amino-terminal domains (disulfide and desmosine cross-links),
r-trimer associations in the interrupted helical domain (with disulfide and desmosine
C;Function:
A;Description: structural component of extracellular basement membrane
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplicat
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F;29-162/Domain: amino-terminal nonhelical, 7S <7SD>
F;163-1440/Domain: interrupted helical <COL>
F;414-452/Region: integrin binding #status experimental
F;597-599/Region: cell attachment (R-G-D) motif
F;917-919/Region: cell attachment (R-G-D) motif
F;968-970/Region: cell attachment (R-G-D) motif
F;1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1451-1551/Domain: collagen alpha 1(IV) chain #status predicted <CT1>
F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;27/Modified site: blocked amino end (Ala) (in mature form) #status experimental
F;31,36,39,41,125,434,467,470/Disulfide bonds: interchain #status predicted
F;45,48,78,90,129,156,172,217,228,231,277,295,298,322,343,361,460,463,497,527,540,543
1082,1084,1099,1117,1132,1150,1165,1182,1185,1188,1206,1235,1265,1283,1304,1319,1328,
F;45,48,78,90,129,156,217,228,231,277,295,298,322,343,361,460,463,497,527,543,573,582
99,1117,1132,1150,1165,1182,1185,1188,1206,1235,1265,1283,1304,1319,1328,1340,1356,13
F;54,63,75,84,87,96,102,105,108,111,117,120,123,138,141,147,150,153,159,167,178,181,1
,419,422,425,439,445,448,451,479,485,491,494,503,512,518,524,530,546,549,552,555,561,
9,745,748,751,754,763/Modified site: 4-hydroxyproline (Pro) #status experimental
F;126/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;129/Modified site: allysine (Lys) #status predicted
F;172,540,947/Modified site: 5-hydroxylysine (Lys) #status atypical
F;272,645,839/Modified site: 4-hydroxyproline (Pro) #status atypical
F;446-447/Cleavage site: Gly-Ile (gelatinase B) #status predicted
F;766,775,784,787,790,796,799,804,810,816,822,834,860,863,869,872,875,887,890,893,899
23,1129,1138,1141,1159,1171,1176,1179,1194,1200,1203,1215,1224,1227,1244,1247,1250,12
431,1437/Modified site: 4-hydroxyproline (Pro) #status experimental
F;1120,1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
F;1120,1268/Binding site: carbohydrate (Lys) (covalent) (partial) #status absent
F;1214,1424/Modified site: 3-hydroxyproline (Pro) #status absent

```
F;1392,1395,1398,1404/Modified site: 4-hydroxyproline (Pro) #status experimental
F;1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
F;1505-1511,1616-1622/Disulfide bonds: #status predicted
F;1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match        16.1%;  Score 609.5;  DB 1;  Length 1669;
Best Local Similarity  28.8%;  Pred. No. 2.4e-21;
Matches  265;  Conservative  63;  Mismatches  296;  Indels  295;  Gaps  54;

Qy    2 GVPGAI----PGGVPGGVPYPGA-----------GLGALG------ 31
          | |::   : |!|| :||!:          ! !|  !
Db  574 GSPGSVGLKGERGPPGSVGFPGSSRGDGPGPPGVGPAGPIGDKGQAGFPGGPGSPGLPG 633

Qy   32 PGGKPLK--PVPPGGLAGAGLGAGLGAFPAVT---PPGALVPGGVADAAAAYKAAKAGAG 85
          !!|!|||   !! |!  |!!  !||            :!|
Db  634 PKGSPGKIVPLPGPPGAEGL-PGSPGFPGPQGDRGFPG--TPGRPGLPGEKGAVGQPGIG 690

Qy   86 LGGVPGVGGLGVSAGANVVPQFPGAGVXPG---------KVPGVGLPGV--TPGGVL 129
          !!|!| : !: : |  |  !||||               !||!|!!||| :  :||||
Db  691 TPGPPGPKGVDGLPGDMGP-PGTPGSPGPENGLPGNPGVGQGKGEPGVGLPGLKGLPG--L 747

Qy  130 PGARFPG------VGVLPGVPTGAGV----------KPKAPGVGAFA--GIPGVGPP 169
          | |!||!      !|| ||!    |           :!|!!|!      !!||!| |
Db  748 PG--TPGTPGEKGSIGV-PGVPGEHGAIGPPGLQGIRGEPGPPGGLPGSVGSSPGVPGIGPP 804

Qy  170 GGPQPGVPLGYPIKAPKLPPGGVGLPYTTGKLPYGYGPGGVVAGAAGKGCYPIGHGVSPCGAA 229
          | |  |! ||                         |!| |
Db  805 GARGP---------------------PGGQGPPGLS------GPPGIKGEKGFPGFFPGILDMFGP-- 841

Qy  230 AAAAKAAAKPGAGRAGVLPGV-GGAGVPGVPG--AlPGIGGIAG-------VGTPAAAA 279
             |!|  |!|!  !  !|!||  ! !!:|!||   :! || !|!         :|| !
Db  842 ---KGDKGAQG-LPGITGQSGLPGLPGQQGAPGIPGFPGSKGEMGVAGI------ 886

Qy  280 AAAAAAKAAKYGAAAAGLVPGGPGFGPGVVGVPGA------VGPGV-------GVPG----- 322
             |! !|! !| |!  |!!|| | !||!  ||!|        !|!|        !||
Db  887 ----PGQPG--SPGPVGAPGLLPGEKGDHGPPGSSGPERGDPGLKGDKG 927

Qy  323 -AGIPVVPGA----------------------------GIPGAAVPGVVSPEAAAKA 350
           ! || ||!                            !||!| !|||!   : :||
Db  928 DVGLPGKPGSMDCKVDMGSMKGQKGDCQGEKGQIGPIGEKGSRGDPG--TPGVPGCKDG---Q 982

Qy  351 AAKAAKYGARPGVCVGGIPTYGVGAGGFPG----FGVGVGGIPGVAGVPSVGGVPGVGG 405
            |! |!| | |!  !|!|!  | | ||         |!| !||!  |!|!||!||! |
Db  983 AGQPGQPGPKGPCDGISGTP----GAPGLPGPKGSVGGMGLPGTPGEKGVP---GIPGPQG 1035

Qy  406 VPGVGISPEAQGAQAAAX---AAKYGVGTPAAAAKAAAKAACFGLVPG------VGYAPG 456
          !||!|  |      ||    |!|!| !|!    !|! !|  !||!|||       !| ||
Db 1036 SPGL--PGDKGKAGEKGQAGPPGIGIPGLRGEKGDQGIAGFPGSPGEKGEKGSIGI-PG 1091

Qy  457 VGVAPGVGVAPG-VGL----------APGVGVARPGVGVAPGVGVAPG-IGHGGVAA 500
          ! |!||   ||| !||           ||!|!!! |!| !|!|!! || !| |!!||
Db 1092 MPGSPGLKGSPGSVGTPGSSPGLPGEKGDKGLPGLDGIPGVKGEAGLPGTPGPTGPAGQKG 1151

Qy  501 AAKSAAKVAAKAQLRAAAGL-GAGIPGLGVG----------VGVPGLGVGAGVPG----- 544
            || !!|  !|  ! ||| | | !!||! |          !|!|||!|!| !||
Db 1152 EPGSDG-IPGSAGDKGEPGLPGRGPGFPGAKGDKGSKGEVGFPGLAGSPGIFGSKGEDG 1210

Qy  545 -LGVGA----GVPGFGAGADDGVRRSLSPELRGDPSSQHLPSTPSSPRVFG--ALAA 596
           :!!|     |!|| | |  |!  !!     !|!  : !! |!|!| |!|    ||!
Db 1211 FMGPGPQGPQGPGLPGSPGHATGGPKGDRGPGQGQQGQGLPG----LPGGMGPGLFGLPGIDGVKG 1266

Qy  597 AKAAKYGAAVPGVLGGLGALGVGVGVPGGVGAGPAAAAAKAAAXAAQFGLVGAAGLGG 656
           |!!||! !!||!!|!!|!|!||!|!||| !| |!!   !|! !!  |!!|!| |!||
Db 1267 DKGNPGWPGAPGVPGKGGDPGFQGMPG--IGGSPGITGSKGDMGPPGVP-GFQGPKGLPG 1323

Qy  657 L-GV---GGLGVPGVGGLGGGIP-----------PARAAAAKYGAAGLGGVLGGGAGGF 699
          | |!   |!!|!|| !|!||!|            |!!  |!!| ||!|||! || |||!
Db 1324 LQGIKGDQGDQGVPGAXGLPGPGPPGPSPVDIIKGEPGLPGEGPGLKGLQGLPGPKGKGQQ 1383

Qy  700 PLGGVAARPGFGLSPIFPG 718
          |!||! |!|!|!  ! !||
Db 1384 GVTGLVGIPG--PPGIPG 1399
```

FT   MOD_RES   448   448   OXIDATIVE DEAMINATION.
FT   MOD_RES   489   489   OXIDATIVE DEAMINATION.
FT   MOD_RES   493   493   OXIDATIVE DEAMINATION.
FT   MOD_RES   544   544   OXIDATIVE DEAMINATION.
FT   MOD_RES   548   548   OXIDATIVE DEAMINATION.
FT   MOD_RES   552   552   OXIDATIVE DEAMINATION.
FT   MOD_RES   606   606   OXIDATIVE DEAMINATION.
FT   MOD_RES   609   609   OXIDATIVE DEAMINATION.
FT   MOD_RES   645   645   OXIDATIVE DEAMINATION.
FT   MOD_RES   649   649   OXIDATIVE DEAMINATION.
FT   MOD_RES   685   685   OXIDATIVE DEAMINATION.
FT   MOD_RES   688   688   OXIDATIVE DEAMINATION.
FT   VARSPLIC   226   239   MISSING (IN ELASTIN B).
FT   VARSPLIC   226   259   MISSING (IN ELASTIN C).
FT   CONFLICT   12   12   RS -> AG (IN REF. 2 AND 3).
FT   CONFLICT   12   12   F -> G (IN REF. 2 AND 3).
SQ   SEQUENCE   747 AA;   64229 MW;   633C03F411643D83 CRC64;

Query Match   65.5%;   Score 2479;   DB 1;   Length 747;
Best Local Similarity   66.8%;   Pred. No. 2.5e-91;
Matches 528;   Conservative 35;   Mismatches 99;   Indels 128;   Gaps 22;

Qy   1   GGVPGAIPGGVFPGGVFVFPGAGLGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT   60
             ||||||||||||||||:||||||| : || : ||| : || : ||| |: |||
Db   27   GGVPGAVPGGVPGGVFPGAGLGGLGVGGLGPGVKPAKPGVGGLVGPGCLGAEGSALPG-A   85

Qy   61   FPGAL--VPGGVADAAAAYK-AAKAGA---GLGGVPGGVGLGVSAGAVVPQ----PGAGV   110
             ||| |||
Db   86   FPGGFFGAGGGAAGGAGAAAYKAAAKAGAAGLGGVGGIGVGVSTGAVVPQLGAGVGAGV   145

Qy   111   KPGKVPGVGLPGVVPGGVLP--GARFPGVGLPGVTPGAGVKPRAPGVGGAFAGIPGVGP   168
             ||||||||||| |||||| : |
Db   146   KPGKVPGVGLPGVVPGGVLPGGAGRFPGIGVLPGVTPGAGVKPKAQVGGAGPGIPGVGP   205

Qy   169   FGGPGPGVPLGTPIKAPKLPGGYGLPTTGKLPVGYGPGGVGAGGAGRAGYPTGVGVGVGZGQA   228
             ||||:|||
Db   206   FGGGQPGLGPLGYPIKAPKLPAGYGLPYKTSKLPYGYGFGVGAGVGAGSAGKAGIPTGVGTGVGPQ-   264

Qy   229   AAAAAAAKPGRGAAGVLP--GVGGAGVPGVGALPGGIGVAGVGTPAAAAAAAAAK   286
             |||||||| |||||| | ||||||||| |||||| :||||:|||||| | ||||||||||
Db   265   AAAAAAAAAAKLGAGGAGVLRGVGRGVGGVGGPGSIPGAGVIPGIGGIAGVGAPGAPAAAAAAAAK   324

Qy   287   AAKYGAAAGL----VP--GGPGPG-PGV----VGVGAGVPGVGVPGGAIP--VVPGAGI   333
             ||:||:|||| : | : ||:||| |||
Db   325   AAKFGAAGGLPGVGVSPEAAAAAAAAKYGARPGVGVGVGVGIPTYGVGAGGPPG----   384

Qy   334   PGRAVPGVVSPEAAAAAKAAKYGARPGVGVGIPTTGVGVGGPGPG--------   382
             || |||| |||||| |||
Db   385   PGVGVPGALSPAATAKAAKAAAKFGARGAVGIGGIPTFGLGPGGFPGIGDAAAAPAAAAA   444

Qy   383   ----VGVGGIPGVAGY-----PSVGGYCVGGYPGVGISPEAQRAAAAAKAAKYGVGTPAA   433
             :| | : || : |
Db   445   KAAKIGAGGVGALGGVVPGAPGAIPGLPGLVGGYPGGYI-----------------PAA   485

Qy   434   AAAKAAKRAAQFGLVPGVGVGVAPGVGVTAPGVGVGLAPGVGVAPGVGVAPGVGVAPGI   493
             ||||||||||||||||||||| : || : || |||
Db   486   AAAKAAAKAAKANQFGLGPGVGVGVPGVVPGVGVAPGVGVAPGIGLGPG-----------GY   534

Qy   494   GPGGVAAAAKSAAKTAAKAOLRAAAGIGAGIPGLGVGGVGVGGVGVGACVPGLGYGAGVPG   553
             ||||||||| :|||| ||||||||||||||||
Db   535   IGAGVPAAAKSAAKAAAKAQFRAAAGLIPAGVPGLGVGAGVPGGLGVGACVPG   594

Qy   554   FGAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGALARAAKRAAKYGAAVPGVLGGL   613
             |||||||||| : ||| : |||
Db   595   PGA---   621

Qy   614   GALGGVGIPGGGVVGAGRAGGARRRGLG-VPGGVGGLGG   672
             || |:||||||||||||| | |||||: |
Db   622   GDLGGAGLPGGVAGVVP-AAAAAKAAAAKAQFGL--GGVGGLGVGGLGAPGGNVGLGG   677

Qy   673   IPPAAAAKAAKYGAAGLGGGVLGGAQQPPLGGVAARPGFGLSPI-------FPGGAC   721
             : ||||||||||||| |||:||
Db   678   VSPAAAAKAAKPGAAGLGGVGLGAGQQPFPIGGGRGGLGVGGKPKPFGGALGALGFPGGAC   737

Qy   722   LGKACGKKRK   731
             ||:|:|||||
Db   738   LGKSCGKKRK   747

RESULT   3

ELS_MOUSE

ID   ELS_MOUSE   STANDARD;   PRT;   860 AA.
AC   P54320;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN).
GN   ELN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Lung;
RX   MEDLINE=95130069; PubMed=7829060;
RA   Wydner K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT   "Use of an intron polymorphism to localize the tropoelastin gene to
RT   mouse chromosome 5 in a region of linkage conservation with human
RT   chromosome 7.";
RL   Genomics 23:125-131(1994).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC   NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC   INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U08210; AAA80155.1; -.
DR   HSSP; P04002; 1WFA.
DR   MGD; MGI:95317; Eln.
KW   Structural protein; Repeat; Signal; Connective tissue.
FT   SIGNAL   1   27   POTENTIAL.
FT   CHAIN   28   860   ELASTIN.
SQ   SEQUENCE   860 AA;   71955 MW;   0CCOBE5AAE1EDD7F1 CRC64;

Query Match   63.1%;   Score 2387;   DB 1;   Length 860;
Best Local Similarity   61.1%;   Pred. No. 1.1e-87;
Matches 546;   Conservative 25;   Mismatches 96;   Indels 226;   Gaps 34;

Qy   1   GGVPGAIPGGVPGGVFVFPGAGLGAL--GGGALGPGGKPLKP-----VPGGLAGAG   48
             ||||||||||||||||||||||:||| |||||||||||:| ||| |||
Db   32   GNVPGGLPGGVPGGVFYIPGAGIGLGGGVGGVGGGGGGALGPGGKPPKPGAGLLGTGAGPGGLGGAG   91

Qy   49   LGAGLGAAFPAVTFP--GALVPGGVADAAAANY-AAKAGAGLGGVPGV----   92
             |||||||| |: ||||||||| |||| :|||:|||||| ||||
Db   92   FGAGLGAFPAGTFPGAGAALPGGALPGGGARGAAAAYKAAAKAGAGLGGVPGGVGVGGVPGGV   151

Qy   93   ---------------GGLGVSAGAVVPGPGAGV---KPGKVPGVGLPGVY   124
             ||||||||||||||||||| : |||:|||||||||||:
Db   152   GVGGVPGGVGVGGVPGGVGVGGIGGLGGLGVGVSGAVVPQVGAGIGAGIAGGKGKVPGVPGVY   211

Qy   125   PGGVLP--GARFPGVGVLPGVPTGAGVKPKPKAPGVGGAPAGIPGVGPGPGPGGLGGVGGYPI   182
             |||| | |||||||||||||||||||||| :|||||||||||||||:
Db   212   PGGVLPGIGARFPGVGVLPGVPGVPTGTGVKAKPGGGGAPSGIPGVGPGVGPGVLPGTPI   271

```
Qy   183  KAPKLPEGIGLPYTTGKLPYTGPGPGGVGAGGAGGKAGYPTGTGVGPQAAAAAAAAKFGA  242
          |||||||||||| |||||||||||  |||||||||||||||||||||||  |||:||
Db   272  KAPKLPGGYGLPYTNGKLPY-----GVRAGAGGKAGYPTGTGVGSQAAAAAA--KAAKYGA  324

Qy   243  GAAGVLPGVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAKAAKYGAAAGLVPGGPG  302
          | |||||||| |:| ||||||||||||||||||||||||||||||| ||||||||
Db   325  GGAGVLPGVGGGGIPGAGAIPGIGIAGAGTPAAAAAAAAKAAKAAKYGAAGGLVPGGPG  384

Qy   303  FGPGVGVPGAGVPGV-GVPG-AGIPVVPGAGIPGAV---PGVVSPAAAKAAKAAKY  357
          |||||:||| :|:|| |||| ||:|||||||||||      |||||||||||||||
Db   385  -----YRLFGKGIPGVGVGGIPGVGGPGIGGPGIVGPGAVSPRAAAKAAKAAKY  439

Qy   358  GAREGVGVGGIPTYGVGAGGFPGFGVGVGIPGVAGVPSVGGVPGVGGVPGVGISPEAQA  417
          ||||:||||||||||||||||:|:|::|:|                  |:||  |:||
Db   440  GARGGV---GIPTYGVAGGFPGYGVNGAGA---------GLGGA-----SP-AAA  476

Qy   418  AAAKAAKY--------------------GVGTPAAAAAKAAAK-A  442
          |||||||||                     ||||||| ||||  |
Db   477  AAAKAAKYGAGGAGALGAGGAGLGGLVPGAVPAVPGAGVPGAGTPAAAAAAAAKAA  536

Qy   443  AQPGLVPGVGAPGVGVAPGVGVAP--GVGLAPGVGVAPGVGVAPGVGVAP-GIGPGGV  498
          |:/ |||/ ||:|| ||||||||| |  |||  |||||||||  :/|  :::  :
Db   537  AKAGLGPGVGGVPG-GV--GVGGIPGGVGVGVPG-GVGPG-GVT-GIGAGPGGLGGAGS  590

Qy   499  AAAAKSAAKVAAKAQLRAAAGLAGGIPGLGVGVGVPGLGVGAGVPGFGGVGVPGFGAGA  558
          |||||||||||||||||||||||||||| || | ||||:|||||||||||||||||||
Db   591  PAAAKSAKAAAKAQYRAAAGLAGAGVPGFGGAGAGVPGFGAGAGVPGAGVPGFGAGA  650

Qy   559  DGGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAAKYGAA---------VPG  608
          | | ::|      |   |||:|              |||||||||||           :|
Db   651  --I---GV---PGPGAG-----------------AVPGSLAANSKAAKYGAAGGLGCPGCGLGGPG  688

Qy   609  VLGGLGALGGVGIPGGVGAG-PAAAAAAKAAKAAAKAAQFGLVGAAGL  654
          ||:|:| :| |: |||||||  |||||||||||||||||:| |||||
Db   689  GLGGPGLGLGAGVPGKVAGNAPPAAAAAAKAAAKAAQYGLGGAGGLGAGGGLGAGLGAGA  748

Qy   655  --------GGLGVGGLGVPGVGGLGGCIPPAAAAAAKYGAAGLGGVLGGAGQQPFLGGVA  705
          |||: |||:||||:||||||    ||||||||||||||:|| ||||
Db   749  GLGAGGAGGIGAGGLGAGGLGAGGLGAGGVSPAAAAKAAKYGAAGLGAGGVL-GARPPPGGGVA  807

Qy   706  ARPGFGLSPIPF-----  731
          ||||||||||||
Db   808  ARPGFGLSPITPGGGGAGGLGVGGKPKPKPYGGALGALGYGQGGCFGKSCGRKRK  860

Qy        GGACLGRACGRKRK  731
          |||:|||||||||||
Db        GGACIGRACGRKRK  860
```

RA   Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;
RT   "Elements of the rat tropoelastin gene associated with alternative
RT   splicing.";
RL   Genomics 12:651-658(1992).

```
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- ALTERNATIVE PRODUCTS: THREE DIFFERENT MRNAS HAVE BEEN FOUND THAT
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M60647; AAA42269.1; -.
DR   EMBL; J04035; AAA42268.1; -.
DR   EMBL; M86372; AAA42271.1; -.
DR   EMBL; M86355; AAA42271.1; JOINED.
DR   EMBL; M86363; AAA42271.1; JOINED.
DR   EMBL; M86364; AAA42271.1; JOINED.
DR   EMBL; M86366; AAA42271.1; JOINED.
DR   EMBL; M86371; AAA42271.1; JOINED.
DR   EMBL; M86376; AAA42272.1; -.
DR   EMBL; M86373; AAA42272.1; JOINED.
DR   EMBL; M86375; AAA42272.1; JOINED.
DR   HSSP; P04002; 1WFA.
KW   Structural protein; Connective tissue; Repeat; Signal;
KW   Alternative splicing.
FT   NON_TER       1      1
FT   SIGNAL       <1     21              BY SIMILARITY.
FT   CHAIN        22    864              ELASTIN.
FT   VARIANT     263    307              MISSING (IN CERTAIN CLONES).
FT   VARIANT     308    308              MISSING (IN CERTAIN CLONES).
FT   VARIANT     809    823              MISSING (IN CERTAIN CLONES).
SQ   SEQUENCE   864 AA;  72786 MW;  456894EB09E79FD4 CRC64;
```

```
Query Match             61.1%;  Score 2311;  DB 1;  Length 864;
Best Local Similarity   58.5%;  Pred. No. 1.1e-84;
Matches  531;  Conservative   26;  Mismatches 102;  Indels  248;  Gaps   34;
```

```
Qy     1  GGVPGAIPGGVPGVPFYPGCAGL-GALGGGALAPGGKPLKP--------VPGGLAGAGL  49
          |||:||||||||||:|||||||| ||||||||||||||||         |||||||
Db    30  GGVPGGLPGGVPGVGVYPGAGIGGALGGAGLGGGALPGGKPGCAGLLGAPGAGPGGLGGAGP  89

Qy    50  GAGLGAFPAVTFPGALVPGGVTADAAAYK-AAKAGGGVPGV----------------  92
          |||| | |||||||| |||| | |||||| |:|||| |||:| |
Db    90  GAGLSY--ASRPGGVLVPGGGAGAAAIKAAAKAGAGLGGLGGVPGGVGGGVPGAVGVG  147

Qy    93  -------GGLGVSAGAVVPQPGAGV----KPGKPGVGLPGVYPGGVLP--GARF  134
          |||||||||||||||||||| |||||||||||||||| |:| |||
Db   148  GVPGAVGGIGGIGGLGVGTGAVVPQLGAGVGAGGKPGKVPGVGLPGVYPGGVLPGGYGARF  207

Qy   135  PGVGVLPGVPGVPGAGVKPRAP-GVGGAFAGIPGVGSPFGGPGGPGPGVPLGYPIKAPKLPGGYGL  193
          |||||||||||||| | ||||| ||||||:||||||||||||||||||||||||||||||||||||
Db   208  PGVGVLPGVPGTGVGVKAKVPGGGAGFGSGIPGVGSPFGGSGQQPGVPLGYPIKAPKLPGGYGL  267

Qy   194  PYTGKLPYGVGPGGVGPGGVAGAGKAGYPTGTGVGPQAAAAAAAAKFGAGAAGVLPGVGVGG  253
          |||||||||||||  ||||:| |||:||:|||||||| |||||||||| | | ||||||||||
Db   268  PYTNGKLPY-------GVRAGAGAGGYPTGTGVGSQ-AAVAAAKAAKYIAGGAGGGVLPGVGVGG  321

Qy   254  AGVPGVGAIPGIGGIAGVGTPAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVGVPGA  313
          |:| ||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   322  GGIPGGAGAIPGIGGITGAGTPAAAAAAKAAKAAKYGAAAGLVPGGPG----VRVPGA  376

Qy   314  GVPGVGVPG---------------AGIPVVPGAGIPGAAV---PGVVSPEAAAKAAKAAKY  357
```

```
RESULT 4
ID   ELS_RAT     STANDARD;     PRT;     864 AA.
AC   Q99372;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN) (FRAGMENT).
GN   ELN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91104868; PubMed=1702999;
RA   Pierce R.A., Deak S.B., Stolle C.A., Boyd C.D.;
RT   "Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";
RL   Biochemistry 29:9677-9683(1990).
RN   [2]
RP   SEQUENCE OF 781-864 FROM N.A.
RX   MEDLINE=88330868; PubMed=2971041;
RA   Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;
RT   "Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.";
RL   J. Biol. Chem. 263:13504-13507(1988).
RN   [3]
RP   SEQUENCE OF 264-533 AND 558-864 FROM N.A.
RX   MEDLINE=92241859; PubMed=1572637;
```

```
          :|:|||:||       |||  | |||    :   || ||| |||||||||||
Db   377 GIPGVGIPGVCGIPGVGGIPGVGGIPGVGAGGPGFGVGV-
                                                  |||   | | |   |||||||||||
Qy   358 GARPGVGVGGIPTYGVGAGGPGFGVGV-----------              385
              ||| |||||||||||
Db   437 GARGGV---GIPTYGVGAGGPPGVGVGAGAGLGGASQAAAAAAAAAAKYGAGGAGILGG 493
Qy   386 ---------GGIPGV-AGVP----SVGGVPGVGVPGVGVGISPEAQAAAAAAAKYG 427
             |||||   ||| || |  |||  ||
Db   494 LVPGAVPGALPGAVPGAGLPGAVPGVPGTGGVPGAG----TPAAAA-------- 542
Qy   428 VGTPAAAAAKAAAKAAAAQFGLVPGVGVAPGVGVAPGVGVAPGVGVAPGV 487
              ||||||||| :|: |||   ||  ||| || |: || |:| ||
Db   543 ---AARARAKAAAAQYGLGPGVGGVGPG-GV--GVGGLPG-GVGPG-GVT-GIGTGPGT 592
Qy   488 GVAPG-IGPGGVAAAKSAAKVAKAQLRAAAGLGAGIPGLGVGVGVGVPGLGVGVPGCLG 546
             |: ||  :| : |||||||||::||||  ||||||||||: ||    |:  |:  : ||
Db   593 GLVPGDGLGGAGTPAAAKSAAKAAAKAQYRAAAGLGAGVPGLGVGAGVPGFGAGAG--GFG 650
Qy   547 VGAGVPGFGAGADEGVRSSLSPELRGDPSSSQHLPSTPSSPRVPGAAAAKAAAKYGAA- 605
              ||||||||||||             |:|||::||:|||||||
Db   651 AGAGVPGFGAGA---------------------VPGSLASKAAKYGAAG 679
Qy   606 ---VPGVLGGLGALGGV----------GIPGGVVGAGPAAAAAAAKAAAAKAACFGLVGAA 652
               :|:|| : || |   |||||||||||||:|||||||
Db   680 GLGGPGGLGGPGGLGGPGGLGGPGGFGGPGGLGGVFGGGVAGGAP-AAAAAAAAAQYGLGGAG 738
Qy   653 GL-----------------GGLGVGGLG----VPGVGVGGLGIPPAAAAKAAKYGAAGLGG 691
             ||                 |||| |||||    |||| | |||||||||||||||||||||
Db   739 GLGAGGLGAGGLGAGGLGAGGLGAGGVIPGAVGLGCVSPAAAKAAAKYGAAGLGG 798
Qy   692 VLGGAGQFPLGGVAARPGFGLSPIFP----------------GGACLGK 724
             || |||  |||||||||||||||||||                 ||||||
Db   799 VL-GARPPPGGGVAARPGFGLSPIYPGGGAGGLGVGGKPPKPYGGALGALGTQGGGCPGK 857
Qy   725 ACGRKKK 731
             :|||||
Db   858 SCGRKRK 864
```

RESULT 5
ELS_CHICK
```
ID   ELS_CHICK    STANDARD;    PRT;    750 AA.
AC   P07916;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN) (FRAGMENT).
GN   ELN
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87242320; PubMed=3593675;
RA   Bressan G.M., Argos P., Stanley K.K.;
RT   "Repeating structure of chick tropoelastin revealed by complementary
RT   DNA cloning.";
RL   Biochemistry 26:1497-1503(1987).
RN   [2]
RP   SEQUENCE OF 85-750 FROM N.A.
RX   MEDLINE=88309083; PubMed=2841924;
RA   Baule V.J., Foster J.A.;
RT   "Multiple chick tropoelastin mRNAs.";
RL   Biochem. Biophys. Res. Commun. 154:1054-1060(1988).
RN   [3]
RP   SEQUENCE OF 457-750 FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=87297534; PubMed=3502711;
RA   Tokimitsu I., Tajima S., Nishikawa T., Tajima M., Fukasawa T.;
RT   "Sequence analysis of elastin cDNA from chick aorta and
RT   tissue-specific transcription of the elastin gene in developing chick
RT   embryo.";
RL   Arch Biochem. Biophys. 256:455-461(1987).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MOST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- ALTERNATIVE PRODUCTS: THERE ARE AT LEAST TWO FORMS OF CHICKEN
CC       ELASTIN THAT ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M18633; AAA48761.1; -.
DR   EMBL; M21880; AAA49082.1; -.
DR   EMBL; M15889; AAA49108.1; -.
DR   PIR; A27264; A27264.
DR   PIR; A26601; A26601.
DR   HSSP; P04002; 1WPA.
KW   Structural protein; Connective tissue; Repeat; Signal;
KW   Alternative splicing.
FT   NON_TER         1      1
FT   SIGNAL         <1     24
FT   CHAIN          25    750       ELASTIN.
FT   DOMAIN         83    686       8 X TANDEM REPEATS.
FT   REPEAT         83    127       1.
FT   REPEAT        219    262       2.
FT   REPEAT        263    318       3.
FT   REPEAT        319    393       4.
FT   REPEAT        394    482       5.
FT   REPEAT        483    554       6.
FT   REPEAT        555    619       7.
FT   REPEAT        620    686       8.
FT   MOD_RES        63     63       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES        66     66       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       111    111       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       115    115       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       156    156       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       159    159       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       198    198       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       200    200       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       235    235       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       252    252       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       256    256       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       297    297       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       301    301       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       354    354       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       357    357       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       427    427       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       431    431       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       513    513       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       517    517       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       520    520       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       586    586       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       590    590       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       593    593       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       655    655       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       658    658       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       719    719       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       721    721       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       743    743       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       748    748       OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       750    750       OXIDATIVE DEAMINATION (POTENTIAL).
FT   VARSPLIC      212    212       G -> GLGGFGGQQPGVPLGYPIKAPKLPG (IN
FT                                  EMBRYONIC ISOFORM).
FT   VARSPLIC      501    501       G -> GVGVPGVGVPG (IN EMBRYONIC ISOFORM).
```

```
FT   CONFLICT      536    536       A -> G (IN REF. 3).
FT   CONFLICT      571    571       G -> A (IN REF. 3).
FT   CONFLICT      610    610       P -> A (IN REF. 3).
FT   CONFLICT      654    654       A -> R (IN REF. 3).
FT   CONFLICT      667    667       P -> R (IN REF. 3).
SQ   SEQUENCE      750 AA;   63697 MW;  E57ECD60C6EE556F CRC64;

Query Match             40.8%;   Score 1542.5;  DB 1;  Length 750;
Best Local Similarity   49.9%;   Pred. No. 1.3e-54;
Matches  424;  Conservative  39;  Mismatches 144;  Indels 243;  Gaps  53;

Qy      1  GGVPGAIP-GGVPGGVEYPGAGLGALG---GGALGPGGKPLKPVPGGLAG-----------   46
           !|||||||  |||||||!|!|!||| :!.!  ! || |||||!!!||||||            |!!
Db     25  GGVPGAIPGGGYPGGGFFPGAGVGGLGAGLGAGLGAGGKPLKPGVSGLGGLGPLGLQPGA   84

Qy     47  --AGIGAGLGAFPAVTPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVP   104
             !!||||||||| |||..|!|| |||| !|!|| || !||!!!!||! !: |!|  |  |
Db     85  GVVGGLGAGLGGAFPGAAFPGA-----ASAAMLKAAKAGAGLGGGGIGGLGAGGVGGVP   138

Qy    105  --------QPGAGYV--KFGKVPGVGLPGVTP-GGVLPGA--RFPGVGVLPGVPTGAGVK   150
                    !|!!    !| :|| !|||!! !||! :! ||||| ||||||||||| |! !
Db    139  GGLGVPGVGVQPGVGVGAAGKPPKVPGAGTPGAGPGGGVLPGAGIRPPGVGVLPGVPTGVGTK   198

Qy    151  PKAPGVGGAPAGIPGVGPEGGPQPGVPLGYPIKAPKLPGGYLPYTGKLPYTGYGPGGV-   209
           !|! |!|||||!        !! !     :!!|| |  !!!!|!|||! ||!!|:
Db    199  AKGPG-AGAPAGI---------------PGGYRLPFVN-----GLGPGGIG            228

Qy    210  AGA-AGKAGYPPGTGVGPQAAAAAAAKFGAGAAGVLPGVGAGAVPGVPGAIPGIGG   268
           ||| ||||||| | !!|||||  !||!|! | !|| :!||!| |!|: !!!:!|:
Db    229  AGVLAGKAGYPPGTGVGAQ---AAAKAAAKYG--AGVLPGAG--GIFGVGGVVPGVGV   280

Qy    269  I--AGVGTPAAAAAA--AAAKAAKYGAAAGLVPGGPGFPGFGVPGVPGSAGVPGVG-VPGFAG   324
           |  : !!|||!|||  |||:!!|!|| || ||||! ! !!|||!| || |!|  |!||
Db    281  VPGAGVGGFPAAAAAAAAAAAAAKAAGANTGA--------GVLPGAGGVPGV-VPGVVPFG--   328

Qy    325  IPVVPG-AGIPGGAAVPGVVSPEAAAAANAAAKAAKYGARPGVGVGGGIPPYGVGAGGFPGF-G   382
            ||!||  ||!!!||| !|||!||||||||||||:! || ! !!||!!|||!!||!|||||| |
Db    329  --LVPGVGIPGVA--GVGTPAGAAAAANAAAKYGA----GVPGVGVPGGVIGGVPGVPG       380

Qy    383  V-GVGGIPGVAGPSVGGVPGVPGGVPGV-GISPEAQAAAAANAAAKAAKYGVGTP-AAAAAKAA   439
           | ||||!|| ||!|||!|||!|||!|||  || || !|!|||!!|||!:| | !| |||||||
Db    381  VPGVPGGVPGVGVPGVPGVPGVPGGVPGV---------GVVGGPAAAAAAAAKA            429

Qy    440  AKAAQPGL--VPGVGV--APGVGTAPGVGGVLAPGVGAPGVAPGVGLGAGIPGL----APGVGV-   489
           |||!!|!  |||!||  |||!|!!|||!!||!||!|||| |!!  !:| |!||    ||||||
Db    430  AKAAAFGAGAGRVPGVGVPGVPGVPGVGV-PGVGV-PGVGV-PGVGV-PGVGVPGVPGVGVGGVP   485

Qy    490  ---APGIGPGGY----------------------------------GVGGVP   534
              ||!!||||!                                   |||||
Db    486  GVVGPVGPGVGVGVPGVGVGVPGGVGVPGGLGPGGLAPGVGGLAPGVAPGGLAPGVG   545

Qy    535  GL--GVGAGVPGL-GTGAGVPGF--GAGADEGVRRSLSPELREGDPSSSQHLPSTPSPR   589
           |!  |!|||!||! |!||||||!  !||!!|!: !!!||
Db    546  GLVPGGVGVSGVGLVPGGVGGLGGLAPGVGVPGV-----------------GGPAAA   584

Qy    590  VPGALAAAKAAKYGAAVPGVLGGL-GALGGV-GIPGGVVGGPAPAAAAAAKAAAKAAAAPG   647
           | | |||||!|!!|||||||||| |!|!!| |!||| ||!!: :
Db    585  ---AKAAAKAAKYGAVGGVPGAVPGAGVPGVTPGVGGVSLVPG-               631

Qy    648  LVGAAGLGGLGVGGL---------GVPGVG-GLGGLPPAAKAAAKYGAAGLGG   691
           |!|| ||!|||!!|:         |||||  :!|!
Db    632  -VGVPGCVVLPGAGIPGVGVQVPGAGAKPPKPGVPGAGVPGVGVGIP-----GLGLVPG   680

Qy    692  V-LGGAGGFPLGGVAARPGFGLSPIFPG------------------------GAC   721
           : :!! ||  || | |||:| |!! ||||                       |||
Db    681  LGVGGLGAGGLGAGVGVPGVPGVSPIFPGGVGGQLGGPGGPKPKTYGGALGALGFRGGVGCA   740

Qy    722  LGKACGGRKRK 731
           || :!| !| |
Db    741  QGKYCGRKRK 750
```

```
RESULT   6
FBOH_BOMMO
ID   FBOH_BOMMO      STANDARD;         PRT;    5263 AA.
AC   P05790; Q26379; Q17220;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   FIBROIN HEAVY CHAIN PRECURSOR (FIB-H) (H-FIBROIN).
GN   FIBH.
OS   Bombyx mori (silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20330562; PubMed=10871375;
RA   Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Esnault C.,
RA   Yang T., Jacquet M., Janin J., Duguet M., Perasso R., Li Z.-G.;
RT   "Fine organization of Bombyx mori fibroin heavy chain gene.";
RL   Nucleic Acids Res. 28:2413-2419(2000).
RN   [2]
RP   SEQUENCE OF 1-168 FROM N.A.
RX   MEDLINE=80045039; PubMed=498286;
RA   Tsujimoto Y., Suzuki Y.;
RT   "The DNA sequence of Bombyx mori fibroin gene including the 5'
RT   flanking, mRNA coding, entire intervening and fibroin protein coding
RT   regions.";
RL   Cell 16:591-600(1979).
RN   [3]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=79211211; PubMed=455439;
RA   Tsujimoto Y., Suzuki Y.;
RT   "Structural analysis of the fibroin gene at the 5' end and its
RT   surrounding regions.";
RL   Cell 16:425-436(1979).
RN   [4]
RP   PARTIAL SEQUENCE FROM N.A.
RC   STRAIN=KINSHU X SHOWA;
RX   MEDLINE=89094868; PubMed=3210244;
RA   Mita K., Ichimura S., Zama M., James T.C.;
RT   "Specific codon usage pattern and its implications on the secondary
RT   structure of silk fibroin mRNA.";
RL   J. Mol. Biol. 203:917-925(1988).
RN   [5]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=94365842; PubMed=7916056;
RA   Mita K., Ichimura S., James T.C.;
RT   "Highly repetitive structure and its organization of the silk fibroin
RT   gene.";
RL   J. Mol. Evol. 38:583-592(1994).
RN   [6]
RP   SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
RC   STRAIN=J-139;
RX   MEDLINE=99296390; PubMed=10366732;
RA   Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,
RA   Takagi T., Mizuno S.;
RT   "Determination of the site of disulfide linkage between heavy and
RT   light chains of silk fibroin produced by Bombyx mori.";
RL   Biochim. Biophys. Acta 1432:92-103(1999).
CC   -!- FUNCTION: FORMS THE SILK FILAMENT; A STRONG, INEXTENSIBLE,
CC       INSOLUBLE AND CHEMICALLY INERT FIBRE.
CC   -!- SUBUNIT: FORMED OF TWO CHAINS. HEAVY AND LIGHT, THAT ARE LINKED BY
CC       A DISULFIDE BOND. HEAVY-LIGHT CHAIN ASSEMBLY IS ESSENTIAL FOR THE
CC       EFFICIENT INTRACELLULAR TRANSPORT AND SECRETION OF FIBROIN.
CC   -!- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE POSTERIOR (PSG)
CC       SECTION OF SILK GLANDS.
CC   -!- DOMAIN: COMPOSED OF ANTIPARALLEL BETA SHEETS. THE STRANDS OF THE
CC       BETA SHEETS RUN PARALLEL TO THE FIBER AXIS. LONG STRETCHES OF SILK
CC       FIBROIN ARE COMPOSED OF MICROCRYSTALLINE ARRAYS OF (-GLY-SER-GLY-
CC       ALA-GLY-ALA-)N INTERRUPTED BY REGIONS CONTAINING BULKIER RESIDUES.
CC       THE FIBER IS COMPOSED OF MICROCRYSTALLINE ARRAYS ALTERNATING WITH
CC       AMORPHOUS REGIONS.
```

CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AF226688; AAF76983.1; -.
DR EMBL; V00094; CAA23432.1; -.
DR EMBL; V00097; CAA23433.1; -.
DR EMBL; S74439; AAB31861.1; -.
DR EMBL; X13869; CAA32076.1; -.
DR EMBL; M35378; AAA27839.1; -.
DR EMBL; AB017362; BAA33147.1; -.
DR PIR; S01844; S01844.
SQ Silk; Signal; Repeat.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 5263 FIBROIN HEAVY CHAIN.
FT DOMAIN 149 5206 HIGHLY REPETITIVE.
FT DISULFID 5244 5244 INTERCHAIN (WITH LIGHT CHAIN).
FT DISULFID 5260 5263
FT CONFLICT 10 10 C -> V (IN REF. 2).
SQ SEQUENCE 5263 AA; 391586 MW; 8ER11D3A0A7440E CRC64;

Query Match 19.0%; Score 719; DB 1; Length 5263;
Best Local Similarity 33.6%; Pred. No. 8.9e-22;
Matches 273; Conservative 20; Mismatches 385; Indels 134; Gaps 27;

```
Qy      5 GALPGGVPGGVFPGPGAGLGALGGGALGPG---GKPLKPVPGGLAGAGLGAGLGAFPAVTF  61
            ||  |: |  |  |||  |     |  ||||         ||| :  |  | || |||| |||
Db    154 GAGAGRAANGSGAGAGSGAAGSGAGAGAGSGTGAGAGAGAGTGAGAGAGAGAGY       213

Qy     62 PGALVPGGVADAAAAYKA---AKAGAGLGVPGVG---GLGVSAGAVVPQPGAGVKPGKV  115
            ||| |  |                ||  ||| ||       :| | ||| |:| ||| |
Db    214 -GA---GAGAGNGAGTGAGAGAGAGTGAGAGAGSGAGAGAGA-----GAGAGYGAA     264

Qy    116 PGVGLPGVTPGGVLPGARF----PGVGVLPGVPTGAGVKPKPAPGVGGAFAGIPGVGPPG  170
            ||   |  | | || | |           |  ||| |  |  |   || |  | | | | |
Db    265 SGAGAGNGTGQGVGVGSGAGAGAGSGAAGSGAGAGSRAAGSGAGAGTGAGNGKAGIPTGTGVGPGAAA  323

Qy    171 GPQPGVPLGIFIKAPKLPGGVGLPYTGKLPIGYGPGGVVAGAGKAGIPTGTGVGPOAAA  230
            || |  | | | ||  | |  |||   ||  |  | |  |  | ||  | | | | | |
Db    324 GAASGTGAGYGAGA----GAGYG------GASGAGAGAGAGAGYGAGYGAGAGA  375

Qy    231 AAAAKANAKPGAGA---------GVLPGVG---GAGVPGVPGAIPGIGGIAGVGTPAAAAA  280
            ||      ||:||||         | ||           | | ||||  | ||   |  |||
Db    376 GAGAGAANGSGAGAGSGAGAGYGAGAGAAGSGAASGAGAGSGAGAGAGS  435

Qy    281 AAAAAAKANKYGAAGLVPG-----------------GPGFGPGVGVPGAGV  315
            ||        |  |  |||                  : |  | |  |  ||| |||:
Db    436 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGTGAGSGAGAGAGAGS-GAGAGYGAGA  495

Qy    316 PGVGVPGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAAKAAKYGARPGVGVGGIPTVGYGA  375
              || |   | |  |  |  | |   :|  |  |   |  || | | || |  || |  ||
Db    496 GGGAASGAGAGSGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGYGA-----GAGA  551

Qy    376 GGFPGFGVGVGGIFGVAGVPSVPGGVPGVGGVPGVGVGISPEAQAAAAAAAAKAAKYGVGTPAAAA  435
            ||   |  | |  |    | || |   | | | |          |  |  | | | |  |  ||||
Db    552 GTGAGAGVGTGAGAGSGAASGAGSGAGAGSGAGAGSGAGAGS-GAGAGGSGAGAGSGAGAGSG  610

Qy    436 AKAAAKAAQ-FGLVPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPIG  494
             |   |     |   ||| |  | |    |  |  |   |  | |  |
Db    611 AGSGAGAGSGAGAGSGAGAGYGAGVGAGYGAGYGAGAGSGAASGAGAGAG  670

Qy    495 PG-------------------GVAAAAKSAAKVIAKAQLRAAAAGLG  521
            ||                    |   | || |  | | | |  |
Db    671 AGTGSSGCFGPTVANGGYSRSDGTEYAWSSDFGTGSGAGAGSGAGAGSGAGAGSG  730

Qy    522 AGI-PGLGVGVSVPGLGVAGVPGLGVAGAGYPGFGPGAGADEGVRRSLSPELREGDPSSSQH  580
            |||  |  |  ||||| | ||   | | |     | |||| |  |   | ||:||||
Db    522 AGI-PGLGVGVSVPGLGVAGVPGLGVAGAGYPGFGPGAGADEGVRRSLSPELREGDPSSSQH  580
```

Db   731 AGAGSGAGAGTGA-GVGVGYGA-GVGVGYGA-GTGAGAGA-GYGAGAGSGGAASGAGAGSGAGSGAGSGA-- 785

Qy   581 LPSTPSSPRVTPGALALAKRAKYGAAVPGVLGGLGALGVGVCIPGGVVGAGPAAAAAAAKAA 640
            | | | | | | ||| | | | | | | | | || |||
Db   786 -----------GAGSGAGSGAGAGSGAGAGSGAGSGAGAGSG----AGAGSGAGSGAGA 829

Qy   641 AKAAQPLVGAAGLG-GLGVG---GLGVPGVGGLGGIPPAAAAKAAKYGAAGLG-GVLGG 695
            || | | | | | | | | ||| | |||| | | | | | |||
Db   830 GSGAGVGSGAGSGAGAGVGYGAGAGV-GYGAGAGSGRAASGAGSGAGSGAGSGAGRNGSG 888

Qy   696 AGQPFLAGVAARPGFGLSPIFPGGACLGKACG 727
            || | | | | | | | ||| || |
Db   889 AGAGSGAGSGAGAGSGAGAGSGAGAGSG 920

RESULT 7
CA24_CAEEL
ID CA24_CAEEL STANDARD; PRI; 1758 AA.
AC P17140;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.
GN LET-2 OR CLB-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94012964; PubMed=7691828;
RA Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;
RT "Genetic identification, sequence, and alternative splicing of the
RT Caenorhabditis elegans alpha 2(IV) collagen gene.";
RL J. Cell Biol. 123:255-264(1993).
RN [2]
RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=90008929; PubMed=2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RT genes are located on separate chromosomes.";
RL J. Biol. Chem. 264:17574-17582(1989).
RN [3]
RP VARIANTS.
RX MEDLINE=94320591; PubMed=8045258;
RA Sibley M.H., Graham P.L., von Mende N., Kramer J.M.;
RT "Mutations in the alpha 2(IV) basement membrane collagen gene of
RT Caenorhabditis elegans produce phenotypes of differing severities.";
RL EMBO J. 13:3278-3285(1994).
CC -!- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
CC -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1
CC DOMAINS.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: FORM I IS PREDOMINANT IN EMBRYOS AND FORM
CC II IS PREDOMINANT IN THE LARVAE AND ADULTS.
CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -!- DISEASE: MUTATIONS IN LET-2 ARE GENERALLY EMBRYONIC LETHAL.
CC

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Z22964; CAA80536.1; -.
DR   EMBL; Z22964; CAA80537.1; -.
DR   EMBL; Z05065; AAA27989.1; ALT_SEQ.
DR   EMBL; U22327; AAA64312.1; ALT_SEQ.
DR   PIR; A34476; A34476.
DR   InterPro; IPR000087; -.
DR   InterPro; IPR001442; -.
DR   Pfam; PF01413; C4; 2.
DR   Pfam; PF01391; Collagen; 23.
KW   Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
KW   Alternative splicing; Glycoprotein; Signal.
FT   SIGNAL   1   26   POTENTIAL.
FT   CHAIN   27   1758   COLLAGEN ALPHA 2(IV) CHAIN.
FT   DOMAIN   27   42   7S DOMAIN.
FT   DOMAIN   42   1527   TRIPLE-HELICAL REGION.
FT   DOMAIN   1528   1758   NONHELICAL REGION (NC1).
FT   DISULFID   1546   1635   OR 1632 (BY SIMILARITY).
FT   DISULFID   1579   1632   OR 1635 (BY SIMILARITY).
FT   DISULFID   1591   1597   BY SIMILARITY.
FT   DISULFID   1654   1750   OR 1747 (BY SIMILARITY).
FT   DISULFID   1688   1747   OR 1750 (BY SIMILARITY).
FT   DISULFID   1700   1707   BY SIMILARITY.
FT   CARBOHYD   248   248   O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT   VARSPLIC   229   264   GDLGAVGVPGPQGPPGPREFTGSSSTVGPKGNPGSKGDK -> G
FT   DIGAMGPAGPGPIASTMSKGTIIGPKGEDLGEKGEK (IN
FT   ISOFORM II).
FT   VARIANT   48   48   G -> E (IN MN114; 73% LETHAL).
FT   VARIANT   366   366   A -> T (IN MN126; 100% LETHAL).
FT   VARIANT   570   570   G -> E (IN MN109; 37% LETHAL).
FT   VARIANT   588   588   G -> R (IN MN151; 96% LETHAL).
FT   VARIANT   597   597   G -> R (IN MN152; 50% LETHAL).
FT   VARIANT   690   690   G -> R (IN MN101; 100% LETHAL).
FT   VARIANT   690   690   G -> E (IN MN129; 100% LETHAL).
FT   VARIANT   737   737   G -> R (IN MN143; 100% LETHAL).
FT   VARIANT   877   877   G -> R (IN G30; 90% LETHAL).
FT   VARIANT   904   904   G -> E (IN E1470; 94% LETHAL).
FT   VARIANT   1003   1003   G -> E (IN MN139; 20% LETHAL).
FT   VARIANT   1125   1125   G -> D (IN G25; 2% LETHAL).
FT   VARIANT   1152   1152   G -> D (IN MN147; 7% LETHAL).
FT   VARIANT   1286   1286   G -> D (IN G37 AND B246; 9% LETHAL).
FT   CONFLICT   1682   1682   P -> L (IN REF. 1).
SQ   SEQUENCE   1758 AA; 167750 MW; 97EE3F3DBB2D2AC5 CRC64;

Query Match   17.8%; Score 675.5; DB 1; Length 1758;
Best Local Similarity   30.1%; Pred. No. 2.2e-20;
Matches 284; Conservative 64; Mismatches 323; Indels 273; Gaps 50;

Qy   2 GVPG-AIPGGPVPGGPVPYPG----AGLGALGGGALGPGGKPLKPVPGGLAGAGLG-ACLGA 55
            |:|| |:||| || | | | ::| | |: :| | ||
Db   634 GLPGPSGPVPGPGAPGIPGEKGDAGLPGLSG--KPGGDGLPGLPGNKGEAGYGQPGQPG 690

Qy   56 PPAVTFPGAL----------VPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAG--AV 102
            ::| |||:| |:|| | | ||| :| :| |:||| | ::
Db   691 FPGAKGDGGLPGLPGTPGLQGMPGEPAPBNQVNPAPGGPGLPGTPGLPGTKGEEGYGRPGE 750

Qy   103 VFQPGAGVKPGKVPGVGLPGV----YPGGVLPGAR-FPGVGVLPGVPGGAG--VKPKAP 154
            :||| ||||| ||| | :| :|| | | || |||:||| ||||
Db   751 VGQPGPGLPGKGDSGLPGPGLPGHPG--VPGDKGFGGVPGLPGIPGPGGKGDVGNPGLP 808

Qy   155 GVGGAPAGIPGVGPFGGP----------------QPGVP-------LGTYIKPAKPLPGVGL 193
            |:|: | ||| | ||| | :| |:| | |:|
Db   809 GLNGQ-KGEPGVGVPGQPGSPGPFGPGLKGDAGLPGTPGLPEGQRGFP-GAPGLKGDGL 866

Qy   194 PYTTGKLPYGVGPG--------GVAGAGKAGYPIGVGV------------------ 224
            | : :|||: || | | | | | |:
Db   GRPGLPGVTGL 923

Db   867 PGLSGQPGY---PGEKGDAGLPGVPGERGSPGFPGQDGLPGVPGNKGEDGLPGLPGVTGL 923
            | :|| ||| | || | | | | | | || || | | || || | ||
Qy   225 ---------------GPQAAAAAAAKGAGAAGV-----------LPGVGG---AGVPGV 259
               : || | | | | |||||| | || :| |:|: ||||
Db   924 KGDLGAPQGSGAPGLPGAPGTPGMKGNAGIPGVPGFKGDGGLPGLPLGNGPKGEPGVPGM 983

Qy   260 PG------------AIPGIGGIAGV-----GTPAAAAAAAAAAAAAKYGAAAGIVP 298
            || ||| | | |: |:||||||| | | :|| | ||
Db   984 PGTPGCMKGNGGLPGLPGLFGRDGLSGGVPGCMKGGDGPSNGLPGEKGEAGPAARDGQKGDAG--LP 1041

Qy   299 GGPGF-GP-GVVGVPGAGVPG---VGVPGAGIPVVPG-AGIPGAAVPGVVSPERAAAKAA 351
             | | | | ||| || || | :| | | ||| | || || || | | :| | |
Db   1042 GQPGLRGPGQGPSGLP--GVPGGFKGGETGLPGTVGQPQGQPGEKGLPG--IPG------- 1086

Qy   352 AKAAKYGARPRVGVGGIPFY-GV-GAGGPPG-----------FGVGVGGIPGVA 392
            | |:||| || ||| | ||| || | |:|| || | |
Db   1087 -KAGRQGAPGSPGQDGLPGFPGVPGMKGSSGYPGSQDGLPGRDGLPGVPGQKGDLGQSGQPGLS 1145

Qy   393 GVPSVGGVPGV---------GGVPGV-----GIS---------PGAQAAAAAKAAKYGV 428
            |||:| || | |||:| :| | ||||| | | | :
Db   1146 GAPGLDGQPGVPGIRGDKGQGGLPGIPGDRGMDGYPGQKGENGYPGQPGLPGLPGLGGEKGFA 1205

Qy   429 GTPAAAAKAA-AKAAQFGL--VPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAP 485
            |||||| | | ||:   :| ||| || |||:| || ||| || | :|
Db   1206 GIPGFPGLKGSPGYPGQDGLPGIPGLKGDSGFFGPGQPGQBGLPGLSGEKGMGGLPGMPGQP 1265

Qy   486 GVGVAPGIGPGVGVAAAKSRAKVAAKAQLRAAAGLGA--GIPGLGVGVSVPGLGVGAGVP 543
            | |: |||| || || | || || | | ||| | |:| |||||| |: ||
Db   1266 GQSIAGPVGPG-----------------APGLQCKDGFPGLPGKGESSLGSLPGAP 1306

Qy   544 GLGVAGVPGPG-GAGADEGVR-------RSLSEDL-RSDDPS-----------SSQHLPSTP 585
              | | ||| | | | || | : | :| |
Db   1307 GLKGFSGMFGFPGAKGDLGANGIPGKRGEKGDGLPGVPGRDGQPGIPGLKGEVGAAGLPGQP 1366

Qy   586 SSPRVPG------------ALAAAKAAKYGAAVPGVLGGLGALGGVGIPG-----GVVGA---- 628
            ||| ||| || | | | | | | ||| | :| | ||
Db   1367 GFPGIPGLKGBOGLPGPGKGAGEAGFPGTPGVPGCYAGEKGDGGLPGLPGRDGLPGADGPV 1426

Qy   629 --GPAAAAAAKAAKRAAQFGLVGRAGL-GGLGVGGL-GVPGVGGLGGIPPAPAAAAKAK 683
               |: ||||| ||| ||| | || |||||| | |||| ||| | | | | ||
Db   1427 GFPGGPSGPQNLVEBFGEKGLP-GLPGAPGLRGGKGMGPLDGPPGGNDGPPGLP------GQ 1478

Qy   684 YGAAGLGGVLGGAGQFPLGGVAARPGFGLSPIFPGGACLGRACG 727
            ||||| |||| |:| | ||| |||||| ||| :||||| |:|
Db   1479 RGNDGVPGAPGLSGBKGMGGLPGFPGLDGQPGGPGAPGLPGAPG 1522

RESULT 8
Y208_MYCTU
ID   Y208_MYCTU   STANDARD;   PRT;   1901 AA.
AC   O53553;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.
GN   RV3508 OR MTV023.15.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC   SUBFAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AL022022; CAA17745.1; -.
DR HSSP; P19972; 1KYD.
DR TubercuList; Rv3508.
InterPro; IPR000084; -.
Pfam; PF00934; PE; 1.
FT SIGNAL   1   30   POTENTIAL.
FT CHAIN   31   1901   HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT   RV3508.
SQ SEQUENCE   1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;

Query Match   17.5%; Score 663; DB 1; Length 1901;
Best Local Similarity   33.6%; Pred. No. 7.1e-20;
Matches 270; Conservative 22; Mismatches 324; Indels 188; Gaps 40;

Qy   2 GVPGAIPGGV--PGGVFYPGAGLGALG--GGAIGPGGKPLKPV-------PGGLAGALG 50
           ||| || || || | ||| | | || || | | ||||
Db   119 GANATTPGGAGGAGGLLFGNGGAAGAPGQAGGPGG----PAGLWNGGPGGAGGSGGG 174

Qy   51 AGLGAPPAVTPGALVPGG-VADAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAG 109
           || || || |||| | |||| ||| | | |||||||||||| | |||
Db   175 TG--------GAGGAGGWLPGVGGVGGGTGGAAGGPGGGLWNGGG------GAG 218

Qy   110 VKPGKVPGVGLPGVTPGGVGVPGARFP---GVGVLPGVPTGSAGVKPKPAPGVGGAPAGIPGV 166
           || | ||||||||||| || | | || | | | | | |||||||
Db   219 GVGGAGGGTGGA--GGRAELLFGAGGAGGACTDCG------PGATGGTCGHGGV 264

Qy   167 GPPGGPQPGVPLGYFIKAPKLPGGYGLPYTTGKLPIGYGPGGVAGAGKAGIPTGTG--V 224
           || || || | || |||||| | | | ||||||||||||||||||||
Db   265 GGDGG--------WLAPGGAGGAG--------GGGGAGGAGSDGGALGGTGGTG 302

Qy   225 GPQAAAAAAAKAAAAKFGAAGAGVL------PGVGGAGVPGVPGAIPGIGGIAGVGTPAA- 277
           || ||||||||||||| | | ||| |||||||||||||||||| | |
Db   303 GTVGGAGGAGGRGCRGLLLGAGGCGGLGAGGGGDGVLGGVSGTGGTGGKGGVGGVAGL 362

Qy   278 AAAAAAAKAAKYGAAAGIV-PGGPGPGPGVVGVPGAGVPGAGVPGAGIPVPGAGIPGA 336
           |||||||||| | || | | | | ||||||||||||||||
Db   363 GGAGGAAGQLPSAGAAGAVGVGGVG---CGQGAGGAGAAGADAP-ASTGLTGGTGFAGG 418

Qy   337 AVPGVVSPEAAAKAAAAKAAKYGARPGVGVGIPTVGVGAGGPPGPGV----------GVG 386
           |||| ||||||||||||| | ||| | ||||||||||| |
Db   419 A--GGVGGQGGNAIAG-----GINGSSGAGGTGGQG-GEAGGMGGSGGADNASGIGADGGRAG 470

Qy   387 GIPGVAGVPSVGVPGVGVPGVGVGISPEAQAAAAKAAKYGVGTP--AAAAKAAAKAAQ 444
           ||| || || ||||||| | | | | | |
Db   471 GTVGGNAGRAGGAAGAGTGGTGGV-------VGAAGXAGIGTGGCQGGAGGAGSAGTDATA 523

Qy   445 FGLVPGVGVAPGVVAPGVVAPGVVAPGVVAPGVVAPGIG---PGGVAAA 501
           || | || | | | || | | |
Db   524 TGATGGTGFSGGAGGAGGAGGAGGAGGNPTGVGGTNGSGGQGGTGGAGGGAGGDNP2GIGGT 583

Qy   502 AKSAAKVRAKAQLRAAAGLGAAGIPGLGVGVVGVPGLSVGAGVPGLGVGAGVPGFGAGADEG 561
           || | |||| | | | | || | ||| |||| ||
Db   584 GGTGGKGGA-------GGAGGQGGSSGAG---GTNGSSGAGGTGGQG--GAGGAG--GAGAD-- 631

Qy   562 VRRSLSPELREGDESSGQHLPSTPSSPRVFGALAAAKAAKYGAAVPGVLGGLG--ALGGV 619
           || | | || | | | ||| | | | | | | |
Db   632 --------------NPTGIGANGGTGGTCGGAGAGGAGGAGGAIGTG-CGGAAVGSVGVNACGIGGT 678

Qy   620 GIPGGVVGAGPA-AAAAKAAAKAAOF-------------GLVGAAGLG 655
           | || | || || || ||| || | ||||||
Db   579 GGTGGVGGAGGAGAAAAGSSATGAGPGAGGAGGAGGEGGAGGAGGKG 738

Qy   656 GL-GVGGLGVP---------GVGLGGIPPAAAAKAAKYGAAGLGGVLGGAGOFPLGGVA 705
           || |||| ||| | | || | | | | | || |||| | | ||| ||
Db   739 GTXGGAGGGAGDNPTGAGGAGGTPGGAGRAGGAGGAT-GTGGTGGVVGATGSAGIGGAG 797

Qy   706 ARP-----------GFGLSPIFPGG 719
                        | || || ||||
Db   798 GRGGGDGGDGASGLGLGLSG-FGGG 820

RESULT 9
SPD1_NEPCL
ID SPD1_NEPCL   STANDARD;   PRT;   747 AA.
AC P19837;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneoidea; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RM MEDLINE=90384959; PubMed=2402494;
RX Xu M., Lewis R.V.;
RT "Structure of a protein superfiber: spider dragline silk.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN [2]
RP SEQUENCE OF 653-747 FROM N.A.
RX MEDLINE=94165058; PubMed=8120021;
RA Beckwitt R., Artidiacono S.;
RT "Sequence conservation in the C-terminal region of spider silk
RT proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and
RA Araneus bicenarrarius (Araneidae).";
RL J. Biol. Chem. 269:6661-6663(1994).
CC -!- FUNCTION: THIS SPIDER MAJOR AMPULLATE SILK POSSESSES UNIQUE
CC   CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF
CC   PSEUDOCRISTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED
CC   WITH ELASTIC AMORPHOUS SEGMENTS.
CC -!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; M37137; AAA29380.1; -.
DR EMBL; U03848; AAB60212.1; -.
DR PIR; A36068; A36068.
KW Silk; Repeat.
FT NON_TER   1
FT DOMAIN   1   655   25 X APPROXIMATE TANDEM REPEATS.
FT REPEAT   1   25   1.
FT REPEAT   26   38   2.
FT REPEAT   39   66   3.
FT REPEAT   67   96   4.
FT REPEAT   97   130   5.
FT REPEAT   131   158   6.
FT REPEAT   159   191   7.
FT REPEAT   192   204   8.
FT REPEAT   205   235   9.
FT REPEAT   236   262   10.
FT REPEAT   263   292   11.
FT REPEAT   293   305   12.

Qy   681 AAKYGRAGLGGVLGGAGQPPLGGVAAR 707
           :| | |||| || |:|
Db   628 AGGAGQCQGYGGY--GSGASASASAASR 652

RESULT 10
CA24_ASCSU
ID   CA24_ASCSU   STANDARD;   PRT;   1763 AA.
AC   P27393;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.
OS   Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC   Ascarididae; Ascaris.
OX   NCBI_TaxID=6253;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91340768; PubMed=1714907;
RA   Pettitt J., Kingston I.B.;
RT   "The complete primary structure of a nematode alpha 2(IV) collagen
RT   and the partial structural organization of its gene.";
RL   J. Biol. Chem. 266:16149-16156(1991).
CC   -!- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
CC   -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
CC   TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
CC   INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1
CC   DOMAINS.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II: ARE
CC   PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC   DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC   G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC   CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC   TRIPLE-HELICAL 7S DOMAIN.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC   ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC   THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC   IV COLLAGENS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M67507; AAA18014.1; -.
DR   PIR; S16366; S16366.
DR   InterPro; IPR000087; -.
DR   InterPro; IPR001442; -.
DR   Pfam; PF01413; C4; 2.
DR   Pfam; PF01391; Collagen; 23.
KW   Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
KW   Alternative splicing; Glycoprotein; Signal.
FT   SIGNAL   1   26   POTENTIAL.
FT   CHAIN   27   1763   COLLAGEN ALPHA 2(IV) CHAIN.
FT   DOMAIN   27   42   7S DOMAIN.
FT   DOMAIN   43   1529   TRIPLE-HELICAL REGION.
FT   DOMAIN   1530   1763   NONHELICAL REGION (NC1).
FT   DISULFID   1548   1637   OR 1634 (BY SIMILARITY).
FT   DISULFID   1581   1634   OR 1637 (BY SIMILARITY).
FT   DISULFID   1593   1599   BY SIMILARITY.
FT   DISULFID   1656   1752   OR 1749 (BY SIMILARITY).
FT   DISULFID   1690   1749   OR 1752 (BY SIMILARITY).
FT   DISULFID   1702   1709   BY SIMILARITY.
FT   CARBOHYD   126   126   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   249   249   O-LINKED (GLYCOSAMINOGLYCAN) (IN FORM II)
FT   (POTENTIAL).

FT   REPEAT   306   333   13.
FT   REPEAT   334   360   14.
FT   REPEAT   361   394   15.
FT   REPEAT   395   424   16.
FT   REPEAT   425   458   17.
FT   REPEAT   459   485   18.
FT   REPEAT   486   512   19.
FT   REPEAT   513   525   20.
FT   REPEAT   526   555   21.
FT   REPEAT   556   582   22.
FT   REPEAT   583   612   23.
FT   REPEAT   613   642   24.
FT   REPEAT   643   655   25.
FT   CONFLICT   662   662   V -> L (IN REF. 1).
FT   CONFLICT   672   672   S -> T (IN REF. 1).
FT   NPGLSGCDVLDALEVVSALIQILGSSSIGQVNPGSAGQA
FT   TQIVGQSVYQAL -> IIVFIDVNSSFKLFSRLFLLSRS
FT   (IN REF. 1).
SQ   SEQUENCE   747 AA;   60528 MW;   B5DE44B0D649E012 CRC64;

Query Match   17.5%;   Score 661.5;   DB 1;   Length 747;
Best Local Similarity 35.2%;   Pred. No. 4.4e-20;
Matches 263;   Conservative 25;   Mismatches 320;   Indels 139;   Gaps 30;

Qy   2 GVPGAIPGGVPGGVTFPGAGLGALGGAGIPGLKPVPGGLAGAGLGAGLGAGGAFPAVTF 61
           |||:|:||| :| ||| | |:||:||| | ||
Db   4 GAAAAAAGGAGQG------GYGGLGGQGAGQGGY-----GGLGGQGAGQGAGAAAAAA 51

Qy   62 PGA----------LVPGGVADAAAAYKAKKAGALGGVPGVGGLGVSAGAVVPGPGRAG 109
               | :   :||: :|:| || |:|| |||| | | |: | | ||||
Db   52 GGAGQGGYGGLGSQGAGRGGQAGAAA--AAAGGAGQGGYGGLGSQGAGRGGLGGQ-GAG 108

Qy   110 VKPGKVPGVGLPGVTPGGVLPGARFPGVGVLPGVPTGAGVKPKAPGVGPAGIPGVGPF 169
           :: || |||| | | ||:| | :| :|| | |
Db   109 AAAAAAGGAGQGGYGGLGNQGAGRGGQGAAAAAGGAG------QGGYGGLGSQGAGRGGL 164

Qy   170 GGPGQPGVPLGYFIKAPKLPGGYGLPYTTGKLPYGPGGYVAGAAGKAGIPTGTGVGPQAA 229
           ||| |::| ||: || || | | || || | ||| | | | |||
Db   165 -----GLGGQGAGQGGY--GGYG-----GLGGQGAGRG-GLGGQGA 215

Qy   230 RAAAAKAAAAKFGAGAGNGQGGLGGOGAGQGGAG----------AAAAAA 283
           | |:| ||| || || ||| ||:|| | | |:|| ||||
Db   216 GAAAAAAGGAGQGAGGQGGLGGQGAGGASAAAAGGAGQGGYGGLGSQGAGRGGEGAGAAAA 275

Qy   284 AAKAAKYGAAAGLVPGGPGP---GTVGVPGAGVPGVGVPGAGIPVVPGAGIPGAAVPG 340
           || |||| || | || | |:|: | || | ||| | ||||| | ||:
Db   276 AAGGAGQGGYGGLGGQGAGQGGYGGLGSQGAGRGGLGGQGAGAAAAAGGAGQGGLGGQG 333

Qy   341 VVSPEPAAAAKAAAAKYGARPGVGVGGIPTYGVGAGGFPGFGVGGIPSVAGVPBSVGGV 400
           ||:| : |:| || | ||| :||| | | | |||||||| |:|:||:|||||
Db   334 AGQGAGAAAAAAGGAGQG------GYGGLGSQGAGRGGLGGQGAGRGAGAVAAAAAGG----AGQ 384

Qy   401 PGVGVGPCPVGVISPEAQAAAAKAAKYGV-------GTPAAAAAKAAAK 441
           |||| ||| |||:| | :| | |:| || ||||||| |||||
Db   385 GGYGGLGSQGAGRGGCQGAGRGGLGGQGAGRGYGGLGCQGAGRAAAAAG 444

Qy   442 AAQFGLVPGVPGGVVAPGVGVAPGVGVALRPGVGVAPGVGVAPGVGVGIGPGGVAAA 501
           || | |:||| ||| ||| | | :||: | || | ||||||||| |: ||||:
Db   445 GAGQGGYGGLG---NQGAGRG------GQGAAAAAGGA-CQGGYGGLG-SQGAGRGGQGGRAG 494

Qy   502 AKSAAKVAAKRAOLRAAAGLGAGIPGLGVGVGVPGGLGVGAGVPGLGVGAGVPGFGAGADEG 561
           | :||:| | : || :| | | | ||| ||||||:| ||:||||| |||||| ||:
Db   495 AAAAAAVGAGQE--GIRGQGAGQGGYG-GLGSQGSGRG-GLGGQGAGAAAAAAGGAGQGG 550

Qy   562 VRRSLSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGV-LGGLGALGGVG 620
           |:: : | | | | : | |:|| ||||| ::|| | |||
Db   551 L-----GGQGAGQ------------------------GAGRA-------AAAGVRQGGYGGLGLGSOG 582

Qy   621 IPGTVVGAGPAAAAKAAAAKRAAKAAAKAAQFGLVGAAGLGLGVGGLGVPGVGGLGGIPPAAAAK 680
           ||||||:|:||| :|||||| || | | :
Db   583 AGRGRGQGAGAAAAA--------GGAGQGGYGGLGGQGYGVGRGGLGGQG-----AGRAA 627

```
FT   VARSPLIC    230    266        GEGQPRGPGQPGPGPVPSTGAKGTIIGPEGAPGMKGEK  ->
FT                                 GIIGPAQPGPPGPPGPREFTGSGSIVGPRGHSGDKGVK (IN
FT                                 ISOFORM I)
SQ   SEQUENCE   1763 AA;  168526 MW;  304F528BC06AAE0D CRC64;
```

```
Query Match       17.0%;  Score 644;  DB 1;  Length 1763;
Best Local Similarity   31.2%;  Pred. No. 3.8e-19;
Matches  277;  Conservative  58;  Mismatches 291;  Indels 262;  Gaps  54;
```

```
Qy       2  GVPGAIPG---GVPGGVFYPGA------------------GLGALGG--------   27
            |:|!| ||    ||  |  ||                    |!||  ||
Db     427  GIPGA-PGVQGPPGLAGPGPGAKGEPGRGVDGQSIPGLPGKDGRPGLDGLPGKRGEMGLP  485

Qy      28  GALGPGGKPLKPVPG-------GLAGAGLGAGLGRFPAVIFPGALVPGGVADAAAAKYA   79
            | :|| | ||  ||        ||:|:|| |
Db     486  GVRGFPGPGDSLNGLPGPPGPRGPQGPKGIDGRDGAPGLPGIPGKGPGDRGGTCAPCARGRKG  545

Qy      80  AKAGAGLGGVPGVGG----QPGAGVKPGKVPGV----GLPGV------------------  123
            |!:|! ||      |      ||| || | ||       |!|:|
Db     546  EKGDAGIAGLPGPGQERGLPGPAVGAPGDDGLPGAPGRPGP-PGPPGQDGLPGLPGGK   604

Qy     124  ------------YPGGV---LPGAR--------FPGVGVLPSVPTGAGVK-----------  150
                        ||!|    || |          ||| |!| :||
Db     605  GEPTQLPLRPGPGYPGQKGETGFGFPGRPGGBLPGKPGIVGAPGLPGPGPGPGPSLTGL   664

Qy     151  PKAPGVGGAPAGIPGV-GPPGGPQPGVPLGYP-IKAPK-LPGGVGLPYTTGKL--------  200
            |:!| ||     | ||:| | ||! !!!!||  | :| |||  |!! ||  :
Db     665  PEKPGKDG-IPGLPGKLGEPGYGQPGMP-GLPGCMKGDAGLPGLPGLPGAVGPMGPPVPES  722

Qy     201  ------PYGYGPGGVAGAGKAGIPTGTGV-GPQAAAAAAAAAAKAKFGAGAAGYLPGV-G  252
                  !|!|!| !!||      ||   ||  |                 |! !! :|
Db     723  QLRPGPPGKDGLPGLPGFKGEAGPPGAPGLQGP------------AGLPG-LPGCMKG  766

Qy     253  GAGVPGVPG--AIPGIGGIAGY-GTPAAAAAAAAAAAKAKYGAAAGL----------   296
            |:||!|| :  ! !| || |   !:|: !|               |!||
Db     767  NPGLPGAPGLAGLPGIPGKGIAGKPGLPGLTGAKGEAG-YPGGPGLPGPGKGEPGPGSTIG  825

Qy     297  VPGGGPPG----------PGVVGVPG------AGVPG---VGVPG-AGIPVVPGA----G  332
            ||||||||          |||||||       |||||   |||||
Db     826  PPGPGPPGGLKGKDGIPGAPGLPGLEGQRGLPGVPGQKGEIGLPGLAGAPGFPGAKGEPG   885

Qy     333  IPGAAVPGVVSPEAAAKAAAKAAKYGARPGVGVGGIPTGVGAGGFPG--FGVGVGGIPG   390
            !| | !!!| |          |:|:||!|            ! !:| |     !:|!
Db     886  LPG--LPGKRGPQ---------------GPPGGPQ-----GAPGPPGQKGDELPGLPG   921

Qy     391  YAGVPSVGGVPGVGGVPGVGISPEAQAAAAKAAKYGVGTPAAAAAKAAAKAAQPGLVPG   450
            ||| |  |||  ||  ||     ||                     |!|| !|||
Db     922  VSGMKGDTGLPGVPGGLAG------PPGQPGFPGQ--KGQPGFFGVYAGKKGEA------GL-PG  969

Qy     451  VGVAPGVGVAPGVGVAPGVGLAPGVGVGVGAPGVGVAPGVGVAPGIGPG-------GVAAAK   503
            :! !| | :| | !! !|| |  |  | ! | !! | !|        ||
Db     970  LPGAPGQKGEQGLAGLPGI---PGMKGAPGIPGAPGQDGLPGL-PGVKGDRGFNGLPGEK  1025

Qy     504  SAAKVAAKAQLRAAGLGAGIPGLGVGVGVPGLGVGAGVPGLGVGAGVPGFSGAGDEGVR   563
            ||                                                  |!: |!||
Db    1026  GBPGPAARDGEKGEPGL-PGQPGLRGPQGPPGL--PGLPGLKGDSGQPGYGA--------  1074

Qy     564  RSLSPELRBGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGVLGGLGAGLGGVGIPG   623
            |! | !|                  !| ! |          || | || | |  |!|| ||
Db    1075  ----PGL-----MGEKGLPGLPGKPGRPGA-PGPKGLDGAPGFPGLKGEAGLPGCAPGLPG  1124

Qy     624  --GIVGAGPARAAAAAKAAAKAAAQPGLVGAAGL-GGLGVGGI-----------GVPGVGGL   670
            !  | ||                           |  | |||             |!:!| ||
Db    1125  QDGLPG-----LPGGKGESGFPGGPGLVGPPGLPGPGKMGAPGIRGEKGDAGLPGLPGSRGL   1179

Qy     671  GGIPGAAAKAAKYGARGLGGVLGGAGQFPLGGVAARPGFGLPGIFPG   718
            ||! | |! |!  || | !!|| |    | | !|| |  | ||  ||
Db    1180  DGL-PGQKGEAGFGAGFLPGVPGVPGKGS--AGAPGFPGLKGEPGLPG  1223
```

RESULT 11
CA14_HUMAN

ID   CA14_HUMAN   STANDARD;   PRT; 1669 AA.
AC   P02462;
DT   01-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
GN   COL4A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelecostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89340433; PubMed=2701944;
RA   Soininen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;
RT   "Structural organization of the gene for the alpha 1 chain of human
RT   type IV collagen.";
RL   J. Biol. Chem. 264:13565-13571(1989).
RN   [2]
RP   SEQUENCE OF 46-1257 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=88083584; PubMed=3691802;
RA   Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
RT   "Complete primary structure of the alpha 1-chain of human basement
RT   membrane (type IV) collagen.";
RL   FEBS Lett. 225:188-194(1987).
RN   [3]
RP   SEQUENCE OF 1-943 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=88029471; PubMed=3311751;
RA   Brazel D., Oberbaeumer I., Dieringer H., Babel W., Glanville R.W.,
RA   Deutzmann R., Kuehn K.;
RT   "Completion of the amino acid sequence of the alpha 1 chain of human
RT   basement membrane collagen (type IV), reveals 21 non-triplet
RT   interruptions located within the collagenous domain.";
RL   Eur. J. Biochem. 168:529-536(1987).
RN   [4]
RP   SEQUENCE OF 28-243.
RX   MEDLINE=86004708; PubMed=4043082;
RA   Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;
RT   "Amino acid sequence of the N-terminal aggregation and cross-linking
RT   region (7S domain) of the alpha 1 (IV) chain of human basement
RT   membrane collagen.";
RL   Eur. J. Biochem. 152:213-219(1985).
RN   [5]
RP   SEQUENCE OF 534-1447.
RX   MEDLINE=85003629; PubMed=6434307;
RA   Babel W., Glanville R.W.;
RT   "Structure of human-basement-membrane (type IV) collagen. Complete
RT   amino-acid sequence of a 914-residue-long pepsin fragment from the
RT   alpha 1(IV) chain.";
RL   Eur. J. Biochem. 143:545-556(1984).
RN   [6]
RP   SEQUENCE OF 1256-1669 FROM N.A.
RX   MEDLINE=85207819; PubMed=2581969;
RA   Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
RA   Cheung M.-C., Prockop D.J., Boyd C.D.;
RT   "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
RT   procollagen reveal an unusual homology of amino acid sequences in two
RT   halves of the carboxyl-terminal domain.";
RL   J. Biol. Chem. 260:7681-7687(1985).
RN   [7]
RP   SEQUENCE OF 1259-1669 FROM N.A.
RX   MEDLINE=85216553; PubMed=2582422;
RA   Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
RA   Kefalides N.A., Myers J.C.;
RT   "Restricted homology between human alpha 1 type IV and other
RT   procollagen chains.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
RN   [8]
RP   SEQUENCE OF 1-28 FROM N.A.
RX   MEDLINE=89034231; PubMed=3182844;
RA   Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;

RT   "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT   collagen are divergently encoded on opposite DNA strands and have an
RT   overlapping promoter region.";
RL   J. Biol. Chem. 263:17217-17220(1988).
RN   [9]
RP   SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
RC   TISSUE=Placenta;
RX   MEDLINE=89005112; PubMed=2844531;
RA   Siebold B., Deutzmann R., Kuehn K.;
RT   "The arrangement of intra- and intermolecular disulfide bonds in the
RT   carboxyterminal, non-collagenous aggregation and cross-linking domain
RT   of basement-membrane type IV collagen.";
RL   Eur. J. Biochem. 176:617-624(1988).

CC   -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC   GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC   MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
CC   NIDOGEN.
CC   -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC   ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC   WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC   -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC   DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC   G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC   CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC   TRIPLE-HELICAL 7S DOMAIN.
CC   -!- PTM: LYSINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT ARE HYDROXYLATED IN ALL CASES AND BIND CARBOHYDRATE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC   ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC   THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC   IV COLLAGENS.

DR   EMBL; M26576; AAA53098.1; -.
DR   EMBL; J04217; AAA53098.1; JOINED.
DR   EMBL; M26550; AAA53098.1; JOINED.
DR   EMBL; M26537; AAA53098.1; JOINED.
DR   EMBL; M26538; AAA53098.1; JOINED.
DR   EMBL; M26548; AAA53098.1; JOINED.
DR   EMBL; M26549; AAA53098.1; JOINED.
DR   EMBL; M26551; AAA53098.1; JOINED.
DR   EMBL; M26552; AAA53098.1; JOINED.
DR   EMBL; M26553; AAA53098.1; JOINED.
DR   EMBL; M26554; AAA53098.1; JOINED.
DR   EMBL; M26555; AAA53098.1; JOINED.
DR   EMBL; M26556; AAA53098.1; JOINED.
DR   EMBL; M26557; AAA53098.1; JOINED.
DR   EMBL; M26539; AAA53098.1; JOINED.
DR   EMBL; M26558; AAA53098.1; JOINED.
DR   EMBL; M26559; AAA53098.1; JOINED.
DR   EMBL; M26560; AAA53098.1; JOINED.
DR   EMBL; M26561; AAA53098.1; JOINED.
DR   EMBL; M26562; AAA53098.1; JOINED.
DR   EMBL; M26563; AAA53098.1; JOINED.
DR   EMBL; M26541; AAA53098.1; JOINED.
DR   EMBL; M26564; AAA53098.1; JOINED.

DR   EMBL; M26565; AAA53098.1; JOINED.
DR   EMBL; M26566; AAA53098.1; JOINED.
DR   EMBL; M26567; AAA53098.1; JOINED.
DR   EMBL; M26568; AAA53098.1; JOINED.
DR   EMBL; M26569; AAA53098.1; JOINED.
DR   EMBL; M26570; AAA53098.1; JOINED.
DR   EMBL; M26571; AAA53098.1; JOINED.
DR   EMBL; M26572; AAA53098.1; JOINED.
DR   EMBL; M26573; AAA53098.1; JOINED.
DR   EMBL; M26574; AAA53098.1; JOINED.
DR   EMBL; M26575; AAA53098.1; JOINED.
DR   EMBL; X07076; CAA68698.1; -.
DR   EMBL; X05561; CRA29075.1; -.
DR   EMBL; M10940; AAA52006.1; -.
DR   EMBL; M11315; AAA52042.1; -.
DR   PIR; S16876; CGHU4B.
DR   MIM; 120130; -.
DR   InterPro; IPR000087; -.
DR   InterPro; IPR001442; -.
DR   Pfam; PF01413; C4; 2.
DR   Pfam; PF01391; Collagen; 21.
KW   Extracellular matrix; Connective tissue; Basement membrane;
KW   Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT   SIGNAL   1   27
FT   PROPEP   28   172   AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT   CHAIN   173   1669   COLLAGEN ALPHA 1(IV) CHAIN.
FT   DOMAIN   173   1440   TRIPLE-HELICAL REGION.
FT   DOMAIN   1441   1669   NONHELICAL REGION (NC1).
FT   CARBOHYD   126   126   N-LINKED (GLCNAC...).
FT   DISULFID 1460   1551   OR 1548.
FT   DISULFID 1493   1548   OR 1551.
FT   DISULFID 1505   1511
FT   DISULFID 1570   1665
FT   DISULFID 1604   1662   OR 1662.
FT   DISULFID 1616   1665   OR 1665.
FT   CONFLICT   237   238   SG -> KE (IN REF. 4).
FT   CONFLICT   241   241   G -> K (IN REF. 4).
FT   CONFLICT   319   319   Q -> A (IN REF. 3).
FT   CONFLICT   719   719   N -> D (IN REF. 5).
FT   CONFLICT   837   837   D -> Y (IN REF. 5).
FT   CONFLICT   842   842   K -> P (IN REF. 5).
FT   CONFLICT   896   896   V -> W (IN REF. 2).
FT   CONFLICT   904   904   E -> Q (IN REF. 5).
FT   CONFLICT   914   914   S -> K (IN REF. 5).
FT   CONFLICT   998   998   K -> P (IN REF. 5).
FT   CONFLICT 1010   1010   S -> K (IN REF. 5).
FT   CONFLICT 1012   1012   E -> Q (IN REF. 5).
FT   CONFLICT 1358   1358
SQ   SEQUENCE   1669 AA; 160611 MW; 3BEBA6DFFB9B8A84 CRC64;

Query Match   16.1%; Score 609.5; DB 1; Length 1669;
Best Local Similarity   28.8%; Pred. No. 8.2e-18;
Matches 265; Conservative 63; Mismatches 295; Indels 295; Gaps 54;

Qy   2 GVPGAI----PGGVPGGVPYPGA----------------GLGALG-------GGALG 31
              ||:|| ||||:||:||: :|| :|:| ::
Db   574 GSPGSVGLKGERGPPGGVGFPGSRGDTGPGPCYGPAGPIGDKGQAGPFGGPGSPGLPG 633

Qy   32 PGGKPLK--PVPGGLAGAGLGAGLGAFPAVT----FPGALVPGGVADAAAYKRAKAGAG 85
              |||| | | | | | | :| :|||: | |: |
Db   634 PKGEPGKIVPLPGPGAEGL-PGSPGPGPQGDRGPPG--TPGRPGLPGEKGAVGQPGIG 690

Qy   86 LGGVPGVGGLGVSAGAVVPQPGAGVKPG---------------KVPGVGLPGV--YPGGVL 129
              |||| | | | | | |:| || | | :|||:|
Db   691 FPGPGPGKGVDGLPGDMGP-PGTPGRPGFNGLPGFGVGGCKGKGEPGVGLPGLKGLPG--L 747

Qy   130 PGARPPG------VGTLPGVPTGAGV--------KPKAPGVGGAFA-GIPGVGPF 169
              ||:| | :|| ||| :| |::|| ||:|: :|:|::|
Db   748 PG--IPGTPGEKGSIGV-PGVGEHGAIGPGLQGIRGEPGPPGLPGSVGSSPGVPGIGPP 804

Qy   170 GGPGQPGVPLGYPIKAPKLPGGIGLPYTTGKLPYGYGPGVGVAGAGKAGYPTGTGVGPQAA 229
              |||| | | | || :| |: || :| | : ||

```
Db    805 GARGP--------------PGGQGPPGLS-------GPPGIKGEKGFPGFPPGLDMPGP---  841
Qy    230 AAAAAKAAAKFGAGAAGVLPGV-GGAGVGPVPG--AIPGIGGIAG-------VGTPAAAA    279
               :|:  |||| |::|:|||  ||  :::|:|         |
Db    842 --------KGDKGAQG-LPGITGQSGLPGLPGQQGAPGIPGFPGSKGPMGVMGT------    886
Qy    280 AAAAAKAAKYGAAAGLVPGGPGFGPGSVVGVPGA--------GVPGV-----GVPG-----  322
               |||| ||  |:|| || |||  ::| |||        ::|  |||
Db    887 --------------PGQPG-SPGPVGAPGLPGEKGDRGFPGSSGPRGDPGLKGDKG       927
Qy    323 -AGIPVVPGA--------------------GIPGAAVPGVVVSPERAAKA            350
               :| |:|                    |:|  ||  |||   ::
Db    928 DVGLPGKPGSMDKVDMGSMKGQKQGDQGEKGQIGPIGEKGSRGDPG--TPSVPGKDG---Q  982
Qy    351 AAKAAKYGARPGVGVGGIPTYGVGAGGPPG-----FGVGVGGIPGSVAGVPSGVGGVPCVGG 405
               :|: || |  | :||  ::|  |  ||         || | ||:|| | | |  |
Db    983 AGQPGPGPGPGDGISGGTP---GAPGLPGPKGSVGMGLPGTPGEKGVP---GIPGPQG    1035
Qy    406 VPCVGISPEQAQAAAAK--AAKYGVGTPAAAAAKAAAKAAQFGLVPG--------VGVAPG  456
               :  |:|: |||          ||:|| ||  |  |||  |  ||        |||| :|
Db    1036 SPGL---PGDKGAKGEKGAGPPGIGIPGLRGEKGDQGIAGFPGSPBGEKGSIGI-PG    1091
Qy    457 VGVAPGVGVAPG-VGL---------APGVGVAPGVGVAPGVVAPG-ICPGGVAA         500
               ||||||  ||  :||         ||||| ||||| |||  | ||
Db    1092 MPGSPGLKGSPGSVGYPGSPGLPGEKGDKGLPGLDGIPGVKGEAGLPGTPGTPGEAQGKRG 1151
Qy    501 AAKSAAKVAAKAQLRAAAGL-GAGIPGLVG----------VGVPGLGVGAGVPG-----   544
               || ||: | :||: |||| | |:||          :|:| |||| |||
Db    1152 EPGSDG-IPGSAGEKGEPGLPGPGPPGFPGAKGDKGSKGVGFPGLAGSPGIPGGSKGEQG  1210
Qy    545 -LGVGA---------GVPGFGAGADEGVRRSLSPELREBGDBSSSQHLPSTPSSPRVPG--ALAA 596
               :|  |          ||  : |::|       | :|| |  :|| ||
Db    1211 FNGPGPGPGQGPGLPGSPGHATEGPKGDRGPQGQPGLPG----LPGEMGPPGLPGIDGVKG 1266
Qy    597 AKAAKYGAAVPGVVLGGLGGALGGVGIPGGVVGAGPAAAAAAKAAAKAAQFGLVGAAGLGG  656
               |:|       | :  ||| |||| ||
Db    1267 DKGNPGWPGAPGVPGPKGDDGFQGMPG--IGGSPGIITGSKGDMGPPGVP-GPQGPKGLPG 1323
Qy    657 L-GV----GGLGVPGVGLGGIP--------PAAAAKAAKYGAAGLGGVLGGAGQF        699
               :|     ||:|   ||::|        | :|| |||| :|| ||  | ||:
Db    1324 LQGIKGDQGQGVP2AKGLPGPGPGPGPPGPYDIIKGEPGLPGPEGGPPGLKGLQGLPGPKGQQ 1383
Qy    700 PLGGVAARPGPGLSPIFPG                                            718
               ||  |  |||| |  |
Db    1384 GVTGLVGLPG---PPGIPG                                          1399
```

RT   converting glycine 521 in the collagenous domain to cysteine in an
RT   Alport syndrome patient.";
RL   J. Biol. Chem. 267:12475-12481(1992).
RN   [3]
RP   SEQUENCE OF 85-1685 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=90337990; PubMed=2380186;
RA   Pihlajaniemi T., Pohjolainen E.R., Myers J.C.;
RT   "Complete primary structure of the triple-helical region and the
RT   carboxyl-terminal domain of a new type IV collagen chain, alpha
RT   5(IV).";
RL   J. Biol. Chem. 265:13758-13766(1990).
RN   [4]
RP   SEQUENCE OF 924-1685 FROM N.A.
RX   MEDLINE=91169491; PubMed=2004755;
RA   Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT   "Characterization of the 3' half of the human type IV collagen alpha
RT   5 gene that is affected in the Alport syndrome.";
RL   Genomics 9:1-9(1991).
RN   [5]
RP   SEQUENCE OF 914-1685 FROM N.A.
RX   MEDLINE=90160375; PubMed=1689491;
RA   Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,
RA   Tryggvason K.;
RT   "Identification of a distinct type IV collagen alpha chain with
RT   restricted kidney distribution and assignment of its gene to the
RT   locus of X chromosome-linked Alport syndrome.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
RN   [6]
RP   SEQUENCE OF 1442-1471 FROM N.A.
RX   MEDLINE=90252791; PubMed=2339699;
RA   Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D.,
RA   Sheer D., Solomon E., Pihlajaniemi T.;
RT   "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
RT   to the region of the X chromosome containing the Alport syndrome
RT   locus.";
RL   Am. J. Hum. Genet. 46:1024-1033(1990).
RN   [7]
RP   SEQUENCE OF 1-20 FROM N.A.
RA   Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.,
RA   Marynen P.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM)
RX   MEDLINE=94133540; PubMed=8301933;
RA   Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,
RA   Cassiman J.-J., Marynen P.;
RT   "Differential splicing of COL4A5 mRNA in kidney and white blood
RT   cells: a complex mutation in the COL4A5 gene of an Alport patient
RT   deletes the NC1 domain.";
RL   Kidney Int. 44:1316-1321(1993).
RN   [9]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97338662; PubMed=9195222;
RA   Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT   "The clinical spectrum of type IV collagen mutations.";
RL   Hum. Mutat. 9:477-499(1997).
RN   [10]
RP   VARIANT AS SER-1564.
RX   MEDLINE=91169492; PubMed=1672282;
RA   Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,
RA   Tryggvason K.;
RT   "Single base mutation in alpha 5(IV) collagen chain gene converting a
RT   conserved cysteine to serine in Alport syndrome.";
RL   Genomics 9:10-18(1991).
RN   [11]
RP   VARIANT AS ARG-325.
RX   MEDLINE=92303559; PubMed=1376965;
RA   Knebelmann B., Deschenes G., Gros F., Hors M.-C., Grunfeld J.-P.,
RA   Zhou J., Hertz J.M., Leinonen A., Tryggvason K., Gubler M.-C., Antignac C.;
RT   "Substitution of arginine for glycine 325 in the collagen alpha 5
RT   (IV) chain associated with X-linked Alport syndrome: characterization
RT   of the mutation by direct sequencing of PCR-amplified lymphoblast

RESULT 12
ID   CA54_HUMAN   STANDARD;   PRT;   1685 AA.
AC   P29400; Q16126; Q16006;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   COLLAGEN ALPHA 5(IV) CHAIN PRECURSOR.
GN   COL4A5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94165049; PubMed=8120014;
RA   Zhou J., Leinonen A., Tryggvason K.;
RT   "Structure of the human type IV collagen COL4A5 gene.";
RL   J. Biol. Chem. 269:6608-6614(1994).
RN   [2]
RP   SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RC   TISSUE=Kidney;
RX   MEDLINE=92316923; PubMed=1352287;
RA   Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT   "Complete amino acid sequence of the human alpha 5 (IV) collagen
RT   chain and identification of a single-base mutation in exon 23

RT   cDNA fragments.";
RL   Am. J. Hum. Genet. 51:135-142(1992).
RN   [12]
RP   VARIANT AS GLU-325.
RX   MEDLINE=93244772; PubMed=1363780;
RA   Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,
RA   Rizzoni G., de Marchi M.;
RT   "De novo mutation in the COL4A5 gene converting glycine 325 to
RT   glutamic acid in Alport syndrome.";
RL   Hum. Mol. Genet. 1:127-129(1992).
RN   [13]
RP   VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX   MEDLINE=94010948; PubMed=8406498;
RA   Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
RA   Tryggvason K., Haggsma-Schouten W.A.G., Roodvoets A.P., Rascher W.,
RA   van Oost B.A., Smeets H.J.M.;
RT   "Identification of four novel mutations in the COL4A5 gene of
RT   patients with Alport syndrome.";
RL   Genomics 17:485-489(1993).
RN   [14]
RP   VARIANTS AS E-400;V-406;V-638;A-638;R-653;R-796;R-869;R-872 & C-1241.
RX   MEDLINE=95322976; PubMed=7599631;
RA   Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
RT   "Detection of 12 novel mutations in the collagenous domain of the
RT   COL4A5 gene in Alport syndrome patients.";
RL   Hum. Mutat. 5:197-204(1995).
RN   [15]
RP   VARIANT AS ARG-1649.
RX   MEDLINE=96213750; PubMed=8651292;
RA   Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
RA   Denison J.C., Fain P.R., Gregory M.C.;
RT   "A mutation causing Alport syndrome with tardive hearing loss is
RT   common in the western United States.";
RL   Am. J. Hum. Genet. 58:1157-1165(1996).
RN   [16]
RP   VARIANTS AS.
RX   MEDLINE=96213754; PubMed=8651296;
RA   Renieri A., Bruttini M., Galli L., Zanelli P., Neri T., Rossetti S.,
RA   Turco A., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
RA   Scolari F., Sessa A., Rizzoni G., Tryggvason K., Pignatti P.F.,
RA   Savi M., Ballabio A., de Marchi M.;
RT   "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
RT   exons of the COL4A5 gene.";
RL   Am. J. Hum. Genet. 58:1192-1204(1996).
RN   [17]
RP   VARIANT AS ASP-1498.
RX   MEDLINE=96233932; PubMed=8829632;
RA   Tverskaya S., Bobrynina V., Tsalykova F., Ignatova M.,
RA   Krasnopolskaya X., Evgrafov O.;
RT   "Substitution of A1498D in noncollagen domain of a5(IV) collagen
RT   chain associated with adult-onset X-linked Alport syndrome.";
RL   Hum. Mutat. 7:149-150(1996).
RN   [18]
RP   VARIANT AS GLN-1677.
RX   MEDLINE=97295089; PubMed=9150741;
RA   Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
RT   "Common ancestry of three Ashkenazi-American families with Alport
RT   syndrome and COL4A5 R1677Q.";
RL   Hum. Genet. 99:681-684(1997).
RN   [19]
RP   VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.
RX   MEDLINE=98112435; PubMed=9452056;
RA   Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
RA   Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
RA   Trivelli A., Pinciaroli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;
RT   "Missense mutations in the COL4A5 gene in patients with X-linked
RT   Alport syndrome.";
RL   Hum. Mutat. Suppl. 1:S105-S109(1998).
RN   [20]
RP   VARIANTS AS, AND VARIANTS.
RX   MEDLINE=99140256; PubMed=10094548;
RA   Plant K.E., Green P.M., Vetrie D., Flinter F.A.;
RT   "Detection of mutations in COL4A5 in patients with Alport syndrome.";

RL   Hum. Mutat. 13:124-132(1999).
CC   -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC   GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC   MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
CC   NIDOGEN.
CC   -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC   ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC   WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC   PROBABLY PRODUCED BY ALTERNATIVE SPLICING. ISOFORM 2 IS A LONGER
CC   ISOFORM FOUND IN KIDNEY, IN WHICH 2 EXTRA G-X-X REPEATS INTO THE
CC   TRIPLE-HELIX DOMAIN ARE INTRODUCED.
CC   -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC   DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC   G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC   CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC   TRIPLE-HELICAL 7S DOMAIN.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC   ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC   THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC   IV COLLAGENS.
CC   -!- DISEASE: DEFECTS IN COL4A5 ARE ASSOCIATED WITH X-LINKED ALPORT
CC   SYNDROME (AS). ALPORT SYNDROME IS CHARACTERIZED BY PROGRESSIVE
CC   GLOMERULONEPHRITIS, OFTEN ASSOCIATED WITH HIGH-TONE SENSORINEURAL
CC   DEAFNESS, SPECIFIC EYE ABNORMALITIES (LENTICONOUS AND MACULAR
CC   FLECKS), AND GLOMERULAR BASEMENT MEMBRANE DEFECTS. IN MALES, THE
CC   TYPICAL TIME COURSE FOR THE PROGRESS OF ALPORT SYNDROME IS:
CC   HEMATURIA BY THE AGE OF 5 YEARS, DEAFNESS AND HYPERTENSION IN

```
Query Match             15.8%;   Score 599.5;  DB 1;  Length 1685;
Best Local Similarity   11.3%;   Pred. No.2e-17;
Matches  249; Conservative   49;  Mismatches  271;  Indels  227;  Gaps  46;

Qy      2 GVPGAIPGGVPGGVFPGAGLGALGLGGGALGPGGKP-LKPVPGGLAGLGAGLGAFPAVT 60
          |:|| |  ||:||:  |   ||    |||   |:||   ::: :      :||
Db    585 GLPG--PKGEPGGITFK------GERGPPGNPGLPGLPGNIGPMG-------P----- 622

Qy     61 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLVSAGAVVPQPGAGVKPGKVPG--- 117
          || ||  :   ||   | |:|| | |:|||  |||  ||  :: |:|| ||| :||:
Db    623 -PGVPGPPGPVGE-----KGIGVAGNPGQPGIPGPKGDPGQ1TPG---KPG-LPGNPG 672

Qy    118 ----VGLPGVPGGVLPGARPGGVGVLPGVPTGGAGVKPKAPGVG------GAFAGIPG- 165
          |||| |  |: | || |    |     || || |: : :: :         |||:||
Db    673 RDGDVGLPG-DPG--LPGQ--PG--LPGIPGSKG-EPGIPGIGLPGPGPGKGFPGIPGP 723

Qy    166 ----------VGFPGPOPGVP.GIYPIKAPKLPGVGLPYTTGKLPYGYCPGGVA 210
          :          || ||  |   |  :|   :  | |  |  |   ||
Db    724 PGAPGPPGRIGLEGPPGGPGFPGFPGKGEPGFALPGP--PGPPGLPGFKGAL----GPKGDR 777

Qy    211 GAAGKAGYPTGTGVGPGVADAAAAAKAAAKFGAGAAGVLPGVGGAGVPVPGAIPGIGGIA 270
          |  :|   ||  ||| |::   |      ::  ||  |:||:||:||:||||:|:|||:|| |
Db    778 GFPGPGGPPGRTGLD--------------GLPGPKGDVGPNGQPG-PMGPPGLPGI- 818

Qy    271 GVGTPAAAAAAAAAAKAAKYGAAGLVPGGPGPGSVVGVPGA-GVPG---VGVPGAGI 325
          |:||  :           :|  :||  | | |  || |||:
Db    819 GVGQP-----------------PGPPG1-PGFIGQPGLHGIPGEKGDPGPPGLDV 855

Qy    326 PVVPG----AGIPGAAVPGVVSPEAAKAAAKAAKYGARPGVVCGGIPPTGVGAGGFPG- 380
          |  ||      ||:| | :  |:   | :  :     |  ::|| :   | |    |:|
Db    856 PGPGERGSPGIPGA--PGPIGPPGGSPGLPGKA------------------GASGFPGT 894

Qy    381 ---FG-VGVGGIPGVAGVVPSVGGVPGVGGVPGVISPEAQAAAAAAAKYGVGTPARAAA 436
          :  | | ||   ||  | ::|:   ||| :||   |       :   :|:| | |:|
Db    895 KGEMGWMGPPGPGPGPLGIPGRSGVPGLKGDDGLQGQPGLPGPPGEKGSKGEPGLPGPPGP 954

Qy    437 K------AAAKAAQFGL--VPGVGVAPGVGVGVAPGGVGVAPGVGVLAPGVVAPGVGVG 488
          :       ::  : |     | |   | |   | |  |    |  |   | :||  :|
Db    955 MDPNLLGSKGEKGEPGLPGIPGVSGPKGYGQLPGDPGQPGLSGQPGLPGPPGPGKGNPGLP 1014

Qy    489 VAPG-IGPGVAAAKSAAKVPAKAQLRAAAGLGAGIPGLGVGVPGLGVGAGVPGLGV 547
          | |     : |         | | |          |  |  | |     :|||
```

```
Db  1015 GQPGLIGPPG------------LKGTIG-DNGFPGFPGQGVPGPPG--PSGVPGQPG 1054
              | |  ||          ||| |  |  |     |   | | |  |   |   |
Qy   548 GAGVPG-FGAGADEGVRRSLSPEL--RPGDPSSSQHLPSTPSSSPRVPGA--------LAA 596
                   :  | :     |||  :         |:    : :  :  :
Db  1055 SPGLPGQKGDKGDPGISIGLPGPKGPPG----LPGYPGNPGIKGSVGDPGLPGLPG 1110
              | ||        | || |  |       | | |
Qy   597 AKAAKYGAAVPGVLGGLGALGGVGI----------PGGVVGAGPAAAAAAAKAAKAA 644
                |  | |                      |  | | |
Db  1111 TPGAKGQPGLPGFPGIPGPPGPKGISGPPGNPGLPGEPGVPGGGGHPGQPGPGEKGPKPG 1170
               |:  | |  ||    |  ||:    : ||| ||||  :
Qy   645 QPGLVGAAG-LGGLGVGGLGPGVGGLGGIPPAAAAKAAKYGAAGLGGVLGAGQPPLGG 703
              | :|:|  | | | |              :: || | |  |    | |  ||
Db  1171 QDGIPGPAGQKGEPGQPGFGNFGPGPGLPGL------SGQKGDGGLPGIPGNPG---LPG 1220
                |   |  |  |     |                |  |    | |  ||
Qy   704 VAARPGF-GLSPIFPG 718
                 : |     |
Db  1221 PKGEPGFHG----FPG 1232
```

RESULT 13
ID CAI4_MOUSE   STANDARD;   PRT; 1669 AA.
AC P02463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490;
RA Muthukumaran G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338558;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
RT from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4]
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2579896;
RA Oberbaeumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-234(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV), and alpha 2(V) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89065738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burbelo P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9]
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=379041;
RA Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC   GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC   MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
CC   NIDOGEN.
CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC   ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC   WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC   DOMAIN (NC1) AT THEIR C-TERMINUS. FREQUENT INTERRUPTIONS OF THE
CC   G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC   CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC   TRIPLE-HELICAL 7S DOMAIN.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC   ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC   THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC   IV COLLAGENS.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; J03758; AAA37439.1; -.
DR EMBL; M23333; AAA51625.1; -.
DR EMBL; J04694; AAA50292.1; -.
DR EMBL; X06777; CAA29946.1; -.
DR EMBL; X02201; CAA26132.1; -.
DR EMBL; M15832; AAA37342.1; -.
DR EMBL; M14042; AAA37342.1; -.
DR EMBL; M12879; AAA37343.1; -.
DR EMBL; M13024; -; NOT_ANNOTATED_CDS.
DR EMBL; M13025; -; NOT_ANNOTATED_CDS.
DR EMBL; M13026; AAA37344.1; -.
DR EMBL; M13027; AAA37345.1; -.
DR EMBL; M13043; AAA37346.1; -.
DR EMBL; J04448; AAA37437.1; -.
DR PIR; A33525; CGMS4B.
DR MGD; MGI:88454; Col4a1.
DR InterPro; IPR000087; -.
DR InterPro; IPR001442; -.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
KW Extracellular matrix; Connective tissue; Basement membrane;
```

```
KW  Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT  SIGNAL       1     27
FT  PROPEP      28    172      AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT  CHAIN      173   1669      COLLAGEN ALPHA 1(IV) CHAIN.
FT  DOMAIN     173   1440      TRIPLE-HELICAL REGION.
FT  DOMAIN    1441   1669      NONHELICAL REGION (NC1).
FT  DISULFID  1440   1551      OR 1548 (BY SIMILARITY).
FT  DISULFID  1493   1548      OR 1551 (BY SIMILARITY).
FT  DISULFID  1505   1511      BY SIMILARITY.
FT  DISULFID  1570   1665      OR 1662 (BY SIMILARITY).
FT  DISULFID  1604   1662      OR 1665 (BY SIMILARITY).
FT  DISULFID  1616   1622      BY SIMILARITY.
FT  CARBOHYD   126    126      N-LINKED (GLCNAC...) (POTENTIAL).
FT  CONFLICT    26     26      A -> P (IN REF. 2).
FT  CONFLICT   186    186      S -> L (IN REF. 2).
FT  CONFLICT   319    319      Q -> L (IN REF. 2).
FT  CONFLICT   369    369      Q -> L (IN REF. 2).
FT  CONFLICT   403    403      L -> F (IN REF. 2).
FT  CONFLICT   481    481      P -> L (IN REF. 2).
FT  CONFLICT   493    493      Q -> H (IN REF. 2).
FT  CONFLICT   712    712      S -> I (IN REF. 2).
FT  CONFLICT   813    813      E -> Q (IN REF. 2).
FT  CONFLICT   982    982      O -> E (IN REF. 2).
FT  CONFLICT  1397   1397      V -> S (IN REF. 3).
SQ  SEQUENCE  1669 AA;  160680 MW;  4291689JE5205889 CRC64;

Query Match       15.6%;  Score 589.5;  DB 1;  Length 1669;
Best Local Similarity   29.1%;  Pred No. 5e-17;
Matches  265;  Conservative   64;  Mismatches 304;  Indels   277;  Gaps    56;

Qy     2  GVPGAI---------PGGVPGGVFYPGAGLGALGGGALGPGGKP-LKPV-PGGLAG-AGLGAGLG  54
             : ||:|            ||:|||  ::|| ||::  ::|  ||| |  ||||| ||||||||
Db   574  GSPGSIGLKGENGPPGVGVGPGS-------RGDIGPPGPGVGPIGPVGPGKGQAGZPGGPG  627

Qy    55  AFPAVTFPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQ-------PGAG  109
             |:|:|      |||||||||||||||||  |||    |  ||           |:|
Db   628  S-------PGLPGPKGSEAGKVVPLPGPGAAGLPQSPGPG----------PQGDRGPFPGTP  672

Qy   110  VKPKGKVPG---YPGGVLPGARFPGVGVLPGVPTGCAGVKPKAPGVGGAFAGIP  164
             || :||      |||:|| ||| ||:||:|| ||| :|| |||| |||||:|:||||
Db   673  GZPG-IPGEKGAVGQPGIGFPG--LPGPK--GVPGLPG-EIG---RPGS2GRPG-FNGLP  722

Qy   165  G------VGPFGGP--------QPGVPLGYP----------IKAPKLPGGYGLPYTTGKLPYG  203
             |      || ||          ||| | |          || |||| || |  ||:| :||
Db   723  GNPGPGQGQKGEPGTGLPGLKGGQPGLP-GIPGGTPGEKGSIGGPGVPSGEQGL--TG-IPPG  776

Qy   204  Y-------GPGGVGAAGKAGVPTGCVGPGQAAAAAAAAKFGAGRAGV-------LPGV  251
             |       |||||||||||||  |||  ||  || ||||||||||   |:|:||
Db   777  LQGIRGDPGGVGWGGAGPPGVP--GIGPPGAMGPPGGEGPPGSSGPPGIKGEKGFPGF  833

Qy   252  GGAGVPGVPG------AIPGIGGIAGV-----GFPAAAAAAAAAKAAKYGAAGLVPG  299
             : || :           || :|   :       |:   |||||||||||||||| ||
Db   834  PGLDMPGPKGDKGSQGLPGGVTGQGSGLPGLPGGQOGTPGVFGPVGSKGEMGVMG------TPG  888

Qy   300  GPGEGRGVVGVPGA-----------GVPG-------AGIPVVPGA-------331
             ||:|  || :|      ||||       |||||||
Db   889  QPG-SPGAGTPGLPGEKGDHGLPGSSGPGRGDPGFKGDKGDVGLPGMPGSMEHVDMGSMK  947

Qy   332  -----------------GIPGAAVPGVVSPERAAKAAAKAAKYGARPGVGVGIPT  370
                              ||||||||||||||||||||  ||||||| |||
Db   948  GQKGDWGDERKGQIGFGDKGSRGDPG--TPGVPGKDG---QAGHGPGPGPKGDPGLSGTP-  1001

Qy   371  YGVGAGGFPG-------PGVGVGGIPGVAGYPSVGGVPGVGGVPGVGISPFAQAAAAAK--A  423
             ||||||      ||||||||||||:|||||||| |  ||||||| ||||||||:
Db  1002  ---GSPGLPGPGKGSVGGNGLPGSSPGEKGVP----GIPGSQGVPG---SPGEKGAKGEKGQS  1052

Qy   424  AKYGVGVGTPAAAAAAKAAKAAQFGLVPG------VGVA--PGVGVAPGVGVAPG-------VG  470
             ||||||||||||||||||||||||  ||        ||  ||   || |       ||
Db  1053  GLPGIGIPGRPGDKGDQGLAGPPGGSPGEKGEKGSAGTPGMPGSPGPRGSPGTGIPGHPGSPG  1112

Qy   471  L--------APGVGVGPGVGVARPGVGVAPG-IGPGGVAAAAKSAAKVAAKAQLRAAAGL-  520
             |           |||  :  ||  |:  ||| ::
Db  1113  LPGEKGDKGLPGLLGGVPGGVKGEAGLPGTPGPTGPAGQKGEPGSDG-IPGSAGEKGEQGVP  1171
```

```
Db  1113  LPGEKGDKGLPGLLGGVPGGVKGEAGLPGTPGPTGPAGQKGEPGSDG-IPGSAGEKGEQGVP  1171
Qy   521  GAGIPGLGVGVGVPGLGVGAGVPGLGVGAGVPGVGFGAGADEGYRRSLSPELRRGDPSSSQH  580
             |||:|: : ::|||||||||||||||||  :|||  || ||  :|:: |: ||| :|| :
Db  1172  GRGFPGFFGSKGDKGSKGEVGFPGLAGSPGIPGV------KGEQGFMGPVPGPQGQPG----  1222
Qy   581  LPSTPSSPRVPGALAAKAAAKYGRAVPGVLGGLGALGGVGIPG----------  623
             |  ||  :|  |       |||||| || ||| ||||| ||
Db  1223  LPGPGHP-----VEGPKGDRGPQGQPGLPGHPGPMGPPGPFPGINGPKGDKGNQGWPGGAP  1277
Qy   624  -------GVVGAGPAAAAAAKAAAKAAQFGLVGAAGL-GGLGVGGL---------  662
                    ||: | ||||||||||||||||:|| ||| |||||:
Db  1278  GVPGPKGDPGFQGMPGIGGSPGITGGK-------GDMGLPGVPGFQGQKGLPGLQGVKGDQG  1332
Qy   663  --GVPGCVGGLGGIP---------PAAAAKAKYGAAGLGGVLGGAGQFPLGGVAARP  708
               |||| :| |||             : ||:|:|||| :| || |::: :: | :|
Db  1333  DQGVPGPGPKGLQGPGPPGPGPYDVIKGEPGLFPGBPGPGLKGLQGPPGPKGQQGVTGSVGLP  1392
Qy   709  GFGLSPIFPG  718
             |  :  ||
Db  1393  G---PPGVPG  1399
```

```
RESULT  14
ID   YQ34_MYCTU     STANDARD;     PRT;    778 AA.
AC   P1933;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2634C.
GN   RV2634C OR MTCY441.04C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver K., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares R., Squares S., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z80225; CAB02341.1; -.
DR   HSSP; P04002; 1ATP.
DR   TubercuList; RV2634c; -.
DR   InterPro; IPR000084; -.
DR   InterPro; IPR001899; -.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE  778 AA;  63131 MW;  DAB20FE58E4999E7 CRC64;

Query Match       15.3%;  Score 580;  DB 1;  Length 778;
```

Best Local Similarity   30.1%;   Pred. No. 7e-17;
Matches   240;   Conservative   38;   Mismatches   285;   Indels   234;   Gaps   39;

```
Qy      2  GVPGAIPGGVPGGVTPGKGALGLGALGGGALGPGGKFLPKPVPGGLAGAGLGAGLGAGGAPPAVTP   61
              |   | | ||     |  |    || | | |   ||  |  ||  || ||
Db    121  GANGADGTGAPGG----PGGIL--LGNGNGGSGAFGQP------GGAGGDAGL--------        162

Qy     62  PGALVPGGVTADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKTPGVGLP  121
              ||    | |      | |  | |   ||   |   ||  |  | |  |      |
Db    163  ---IGNGGT-------GGKGGDGLVGSGAAGGVGRGGWLL--GNG-------GTGGA     201

Qy    122  GYPGPGVLPGARPGCVGVLPGVPTGAGVKPKAPGVGGAFAGIPGVGPFGGPGPGVPLGYTP  181
              ||           |  |   | |   | |     | |:  | |||  ||  |   | |
Db    202  GGAAGATLVG----------GTG-------GVGGA-TGLIGSSGGPGG--------        230

Qy    182  IKAPKLPGGVGLPYTTGKLPYGPGVGPGVAGAGKAGYPTGTGVGPQAAAAAAAKAAAKFG   241
              | |       |   | | | |  ||   | |    | |  || || | |    || |||
Db    231  --AGGAAAGVG---TTGSVG--GSGSVGVGVPGNGGPGAGGLG--AAGGVGGAAASVFS    279

Qy    242  AGAAGVLPGVGAGVPGVP3A-----IPGIGGIAGVGPPRAAAAAAAAAKAAKYGAAAGLV  297
              | | |    | | |  |      || |||    |    | || |      ||| |
Db    280  TGGGG------GVGSGAPGGDGAGPLLIGNGGVGGLG------        327

Qy    298  -----PGGPGFGPGVVGPGVGPSVGVPG--------AGIPVVPGAGIPG--AAVPG-V   341
                   || | || |   | || |      ||| || | |     | ||
Db    298  -----PGGPGFGPGVVGPGVGPSVGVPG--------AGIPVVPGAGIPG--AAVPG-V   341

Qy    342  VSPEAAAKAAAKAAK----------------YGARPGVG-VGGIPPYG------------VGAG  376
              |                          || |   |||    || |          ||
Db    386  ANPNTGANNGTDNSGNGNQTGGNNGQPGAGGVEAGGVGQGQCGLGESLLDGNDGTGGKGGAG  445

Qy    377  GFPGPGVGVGGIPGVAGVPSV------------------GGVPGVGGVPGVGISPE       414
              | | | || |  | |  |                  |   | | | ||  ||
Db    446  GTAGTDGGAGGAGGAGGIGETDGSAGGVATGCGGDGATDGGVGGAGKGGQGHNTG       505

Qy    415  AQAAAAAKAAKYGVGTPAAAA-------KAAAKAAQFGLVPGVGVGVAPGVGVGVAP      467
              |                    |           |   | | | | | ||
Db    468  GVGLAPGVVAPGVVAPGVVAPGIPGPGGVAAAAKSAAKVAAKAQLRAAAGLG--AGIP     525

Qy    468  GVGLAPGVVAPGVVAPGVVAPGIPGPGGVAAAAKSAAKVAAKAQLRAAAGLG--AGIP  525
              |     |                                        |  |
Db    506  VGDAFGDGIGDGNGALGAAGGNGGTGGAGGNGGGMLIGNGAGGAGGTGGTGGGG     565

Qy    506  VGDAFGDGIGDGNGALGAAGGNGGTGGAGGNGGGMLIGNGAGGAGGTGGTGGGG      565
              |                                      |  |  || |  ||
Db    556  AAGPAGGVGGAGGEGLIDGAGTAEG-GTGGLGGL--------GGVGGTGGMGGSGGVG    614

Qy    526  GLGVGVG-VPGLGVGAGVPGLGVGAGVPGFGAGADEGVRSLLSPELREGDPSSSQHLPST  584
              |   |       |  |  |   | |         |   |     |  |  |  |
Db    615  GNGGAVGSLIGLGGGGAGGVGGTGGIGGIGGGAGGNG----------------------  651

Qy    585  PSSPRVPGALAAKAAKYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAAAKAAAKAA  644
              || |       | |  | |  |  ||   | | |  |  || | | |
Db    652  ------GAGGAGTTGGGATI----GGGGGTGGVGAGGTGGTGGAGGTTGGSGGAG--   698

Qy    645  QFGLVGAAG-LGGLGVGGLGVPGVGVGLGGIPPAAAAAKAAKYGAAGLGGVGLGAGGPPLGG  703
              |   | | | |    ||    | |  | | |  | |     ||  | | | | | | |
Db    699  --GLIGWAGAAGGTGAGGTG--GGGGLGG-------QGGNGNGGTGATGGGGPFALGG     747

Qy    704  VAARPGFGLSPIFPGGA 720
              |    | |   | | ||
Db    748  NGGAGGAGGS---PGGS 761
```

RESULT 15
CAI1_HUMAN
ID   CAI1_HUMAN   STANDARD;   PRT;   1454 AA.
AC   P02452; Q15176; Q14037;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   COLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN   COL1A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;

RN   [1]
RP   SEQUENCE OF 1-472 FROM N.A.
RX   MEDLINE=89025644; PubMed=3178743;
RA   Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA   Jaenisch R., Prockup D.J.;
RT   "Structure of a full-length cDNA clone for the prepro alpha 1(I)
RT   chain of human type I procollagen.";
RL   Biochem. J. 253:919-922(1988).
RN   [2]
RP   SEQUENCE OF 1-181 FROM N.A.
RX   MEDLINE=84270697; PubMed=6462220;
RA   Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA   Myers J., Williams C., Ramirez F.;
RT   "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT   conservation of a pattern of introns and exons.";
RL   Nature 310:337-340(1984).
RN   [3]
RP   SEQUENCE OF 152-301.
RC   TISSUE=Skin;
RX   MEDLINE=71038625; PubMed=5529814;
RA   Click E.M., Bornstein P.;
RT   "Isolation and characterization of the cyanogen bromide peptides from
RT   the alpha 1 and alpha 2 chains of human skin collagen.";
RL   Biochemistry 9:4699-4706(1970).
RN   [4]
RP   SEQUENCE OF 263-268.
RC   TISSUE=Skin;
RX   MEDLINE=71001508; PubMed=4319110;
RA   Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT   "A comparative study of glycopeptides derived from selected
RT   vertebrate collagens. A possible role of the carbohydrate in fibril
RT   formation.";
RL   J. Biol. Chem. 245:5042-5048(1970).
RN   [5]
RP   SEQUENCE OF 425-1454 FROM N.A.
RX   MEDLINE=84080385; PubMed=6689127;
RA   Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA   Prockop D.J.;
RT   "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT   pro alpha 1 chain of human type I procollagen. Statistical evaluation
RT   of structures that are conserved during evolution.";
RL   Biochemistry 22:5211-5223(1983).
RN   [6]
RP   SEQUENCE OF 1229-1454 FROM N.A.
RC   TISSUE=Bone;
RX   MEDLINE=88124208; PubMed=3340531;
RA   Maekelae J.K., Raassina M., Virta A., Vuorio E.;
RT   "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT   domain.";
RL   Nucleic Acids Res. 16:349-349(1988).
RN   [7]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=88097389; PubMed=3480516;
RA   Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT   "Regulatory elements in the first intron contribute to
RT   transcriptional control of the human alpha 1(I) collagen gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN   [8]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE=85130970; PubMed=2857713;
RA   Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT   "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT   Promoter structure, AluI repeats, and polymorphic transcripts.";
RL   J. Biol. Chem. 260:2315-2320(1985).
RN   [9]
RP   SEQUENCE OF 1-44 FROM N.A.
RX   MEDLINE=88033098; PubMed=2822714;
RA   Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA   de Wet W.J.;
RT   "DNA sequences in the first intron of the human pro-alpha 1(I)
RT   collagen gene enhance transcription.";
RL   J. Biol. Chem. 262:15151-15157(1987).
RN   [10]

RP   REVIEW ON VARIANTS.
RX   MEDLINE=91184577; PubMed=2010058;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in collagen genes: causes of rare and some common diseases
RT   in humans.";
RL   FASEB J. 5:2052-2060(1991).
RN   [11]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX) and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [12]
RP   REVIEW ON OI VARIANTS.
RX   MEDLINE=91374476; PubMed=1895312;
RA   Byers P.H., Wallis G.A., Willing M.C.;
RT   "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL   J. Med. Genet. 28:433-442(1991).
RN   [13]
RP   REVIEW ON OI VARIANTS.
RX   MEDLINE=97169389; PubMed=9016532;
RA   Dalgleish R.;
RT   "The human type I collagen mutation database.";
RL   Nucleic Acids Res. 25:181-187(1997).
RN   [14]
RP   VARIANT OI-II CYS-1166.
RX   MEDLINE=86287390; PubMed=3016737;
RA   Cohn D.H., Byers P.H., Steinmann B., Gelinas R.E.;
RT   "Lethal osteogenesis imperfecta resulting from a single nucleotide
RT   change in one human pro alpha 1(I) collagen allele.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
RN   [15]
RP   VARIANT OI-II ARG-569.
RX   MEDLINE=87222295; PubMed=3108247;
RA   Bateman J.F., Chan D., Walkers I.D., Rogers J.G., Cole W.G.;
RT   "Lethal perinatal osteogenesis imperfecta due to the substitution of
RT   arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RT   collagen.";
RL   J. Biol. Chem. 262:7021-7027(1987).
RN   [16]
RP   VARIANT OI-II CYS-926.
RX   MEDLINE=88033031; PubMed=3667599;
RA   Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT   "A point mutation in a type I procollagen gene converts glycine 748
RT   of the alpha 1 chain to cysteine and destabilizes the triple helix in
RT   a lethal variant of osteogenesis imperfecta.";
RL   J. Biol. Chem. 262:14737-14744(1987).
RN   [17]
RP   VARIANT OI-II ARG-842.
RX   MEDLINE=88298828; PubMed=3403550;
RA   Bateman J.F., Lamande S.R., Dahl H.H., Chan D., Cole W.G.;
RT   "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
RT   chain in lethal perinatal osteogenesis imperfecta. Demonstration of
RT   the peptide defect by in vitro expression of the mutant cDNA.";
RL   J. Biol. Chem. 263:11627-11630(1988).
RN   [18]
RP   VARIANT OI CYS-1195.
RX   MEDLINE=89218628; PubMed=3244312;
RA   Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT   "A cysteine for glycine substitution at position 1017 in an alpha
RT   1(I) chain of type I collagen in a patient with mild dominantly
RT   inherited osteogenesis imperfecta.";
RL   Mol. Biol. Med. 5:197-207(1988).
RN   [19]
RP   VARIANT OI-II VAL-434.
RX   MEDLINE=89255493; PubMed=2470760;
RA   Patterson E., Smiley E., Bonadio J.;
RT   "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
RT   mutation.";
RL   J. Biol. Chem. 264:10083-10087(1989).
RN   [20]

RP   VARIANT OI-IV SER-1010.
RX   MEDLINE=89305591; PubMed=2745420;
RA   Marini J.C., Grange D.K., Gottesman G.S., Lewis M.B., Koeplin D.A.;
RT   "Osteogenesis imperfecta type IV. Detection of a point mutation in
RT   one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
RL   J. Biol. Chem. 264:11893-11900(1989).
RN   [21]
RP   VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.
RX   MEDLINE=89380165; PubMed=2777764;
RA   Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;
RT   "Characterization of point mutations in the collagen COL1A1 and
RT   COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL   J. Biol. Chem. 264:15809-15812(1989).
RN   [22]
RP   VARIANT OI SER-1022.
RX   MEDLINE=90062068; PubMed=2511192;
RA   Pack M., Constantinou C.D., Kalia K., Nielsen K.B., Prockop D.J.;
RT   "Substitution of serine for alpha 1(I)-glycine 844 in a severe
RT   variant of osteogenesis imperfecta minimally destabilizes the triple
RT   helix of type I procollagen. The effects of glycine substitutions on
RT   thermal stability are either position of amino acid specific.";
RL   J. Biol. Chem. 264:19694-19699(1989).
RN   [23]
RP   VARIANT OI-II CYS-1082.
RX   MEDLINE=89109573; PubMed=2913053;
RA   Constantinou C.D., Nielsen K.B., Prockop D.J.;
RT   "A lethal variant of osteogenesis imperfecta has a single base
RT   mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
RT   chain of type I procollagen. The asymptomatic mother has an
RT   unidentified mutation producing an overmodified and unstable type I
RT   procollagen.";
RL   J. Clin. Invest. 83:574-584(1989).
RN   [24]
RP   VARIANT OI CYS-272; CYS-704 AND CYS-896.
RX   MEDLINE=90009313; PubMed=2794057;
RA   Starman B.J., Eyre D., Charbonneau H., Harrylock M., Weis M.A.,
RA   Weiss L., Graham J.M., Byers P.H.;
RT   "Osteogenesis imperfecta. The position of substitution for glycine by
RT   cysteine in the triple helical domain of the pro alpha 1(I) chains of
RT   type I collagen determines the clinical phenotype.";
RL   J. Clin. Invest. 84:1206-1214(1989).
RN   [25]
RP   VARIANT OI-II CYS-422.

```
Query Match          15.3%;  Score 580; DB 1; Length 1464;
Best Local Similarity  33.7%;  Pred. No. 1.1e-16;
Matches 268;  Conservative 32;  Mismatches 342;  Indels 154;  Gaps 47;

Qy      2  GVPGAIPGGVPGGVFYPG-AGL-GALG-GGALGPGGGKPLKP---VPGGLAGAGLGAGLG  54
           |:||  |   ::    |:  |    |: | :||   |: |    : :||   ||| |:||
Db    254  GLPGT--AGLPGMKGHRGFSGLDGAKGDAGPAGPGKGEPGSPGNGAPGGMGPRGLPGERG 311

Qy     55  AFPAVTFPGALVPGGV-----ADAAAYKAAKAGAGLGGVPGVGLGVSAGAVVPGPGAG 109
           |:||  |      | |      : :|    :| |||  | || : | :|| || |: |:
Db    312  R-----PGAPGPAGARGNDGATGAAGPPGTCPAGPGPFPGAVGAKGEAGPGGPRGSEG 365

Qy    110  VKPGKVPG-VGLPGVTPGGVLPGARFPGVGVLPGVTPGAGVKFKAPGVGGAGAGIPGV-G 167
           | |  |    |||  :||   |    | | ||   |:| |  |  || ||  |: |  |
Db    366  --PGGYRGEPGPPG-PAGAAGPAGNPGADGQPGAKGANG---APGIAGA-PGFPGANG 416

Qy    168  PFGGPGPVGPLGYTPIKAPKLPDGGYGLPTTGKLPYGYGPGGVAGAAGAGYFPCTRGV-GP 226
           | |  |  | |  |  :|   |:  ||        :|||| ||  |: |||| || |  ||
Db    417  PSGPQGPGPPG-----PK--GNSGEP-------GAPGSKGDTGAKGEPGPVGVGVQGP 459

Qy    227  QARAAAAARAAKPGAGAAGVLPGV----GGAGVAGPGVGVPGA--TPGIGGIAGV-GTPAAAA 279
           | ||  | |:|| ||   | ||||     ||| ||| || ||   |::| |||  |:|||
Db    460  PGPAGEGKGARGGEPGPTG-LPGPPGERRGGPGPGSRGGFPGADGVAGPGPAGGERGSPCPAG 518

Qy    280  AARAAAKAKYG-----AARGIV-PGGPGPGVGVVPGA----GVPG-VGVPG----A 323
           | |:: | | :     ||| |  :|||  | :  | |     |||  |  |    |
Db    519  PKGSPGEAGRPGEAGLPGAKGLTGSSPGSPG-PDGKTGPPGGPAGDGDRPGPGPPGGAKGGA 577

Qy    324  GIPVPG----AGIPG-AAVPGVVSPEAAAKAAAKAAKYGAARPGVGVGGIPTYGVGAGGP 378
           |||  |     | ||   | ||  :||  ||  |||  | ||| | ||   |  || | |
Db    519  PKGSPGEAGRPGEAGLPGAKGLTGSSPGSPG-PDGKTGPPGGPAGDGDRPGPGPPGGAKGGA 577
```

```
Db      |:    ||    ||||  ||   |  |     |   |     |   :  | :  ||:   | —|—  |    |
    578 GVMGFPGPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQGPGPGAG-PAGERGEQG- 635

Qy  379 PGRGVGGVGIPGVAGYPSVGGVPGVGGVPGVGISPRAQAAAAAKA--AKYGVGTPAAAAA 436
        |—|:|:||| ||::|| |||| | |  ||    :|  ||:   :   ||:|:|:
Db  636 PAGSPGPGGLPGPAGPPGEAGKPGEQGVPGDLGAPGPSGARGERGFPGERGVQGPPGPAG 695

Qy  437 KAAAKLAAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGV-------- 487
        |  | :|||      |||  ||| ||| ||||:  |  ||||  | |:
Db  696 PRGANGA-----PGNDGAKDGAGAPG--APGSQGAPGLQGMPGERGAAGLPGPKGDRGD 747

Qy  488 -----------GVAPGIGPGGVAAAAKSAAKVAAKAQLRAAAGLGA-GIPGLGVG 530
            |:  |||||—|—|:  |    | |—|—|||| |
Db  748 AGPKGADGSPGKDGVRGLLGPIGPPGPPAGPAGPGDKGESGPSG--PAGPTGARGRAPGDRGE 804

Qy  531 VGYPGLGVGAGVPGSLGVGAGVPGF-GAGADEGVRRSLSPELREGDPGSSSGQHLPSTPSSPR 589
        |—|—|| ||||  |  |||  ||    :       |   :|
Db  805 PGPPGPAGFPAGPPG--ADGQPGAKGEPGDRAGAKGDAGP-------PG------PAGPAGP- 849

Qy  590 VPGALA-----AAKAAKYGAAVPGVLGGGIGALGGVGIPGGVVGAGPAAAARAAKAAKAA 644
        ||:     ||| ||||||       |  :|  |||  | | |||||||||||||||||
Db  850 -PGPIGNVGAPGAKGARGSAGPPGATGPTGAAGRVGPGPGSGNAGPPGGPGPPAGKEGGKG 908

Qy  645 QFGLVGAAGLGGLGVGGLGVPGVGGLGGIPPA------AAAKAAKYGAAGLGGVLGGAGQ- 698
        | | |:||||||||| |||| ||| |||       |  |||:||||||||| :|| ||
Db  909 PRGETGPAGRPG-EVGPPGPPGPAGEKGSPGADGPAGAPGTPGPQGIAGQRGYVGLPGQR 967

Qy  699 ----FP-LGGVAARPG 709
        ||| :| ||
Db  968 GERGFPGLPGPSGEPG 983
```

Search completed: April 24, 2001, 16:42:41
Job time: 447 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:36:20 ; Search time 62.39 Seconds
(without alignments)
61.891 Million cell updates/sec

Title: US-09-340-736-2
Perfect score: 1017
Sequence: 1 FPGFGVGVGGIPGVAGVPGV........GVAPGVGVAPGVGVAPAIGP 201

Scoring table: BLOSUM62
GapOp 10.0 , GapExt 0.5

ched: 185757 seqs. 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1017	100.0	201	2	US-08-911-364-2	Sequence 2, Appli
2	700	68.8	731	2	US-08-911-364-1	Sequence 1, Appli
3	662.5	65.1	792	2	US-08-678-039A-40	Sequence 40, Appli
4	401	39.4	988	1	US-08-212-237A-3	Sequence 5, Appli
5	401	39.4	988	1	PCT-US95-02772-5	Sequence 5, Appli
6	397.5	39.1	972	1	US-08-212-237-7	Sequence 7, Appli
7	397.5	39.1	972	1	PCT-US95-02772-7	Sequence 7, Appli
8	392	38.5	1056	5	PCT-US95-02772-6	Sequence 6, Appli
9	392	38.5	1056	5	PCT-US95-02772-6	Sequence 6, Appli
10	391.5	38.5	1024	1	US-08-212-237-6	Sequence 6, Appli
11	391.5	38.5	1024	5	PCT-US95-02772-8	Sequence 8, Appli
12	390	38.3	832	1	US-08-212-237-4	Sequence 4, Appli
13	390	38.3	832	5	PCT-US95-02772-4	Sequence 4, Appli
14	382	37.6	936	5	US-08-212-237-3	Sequence 3, Appli
15	382	37.6	936	5	PCT-US95-02772-3	Sequence 3, Appli
16	379.5	37.3	1413	1	US-08-175-155-39	Sequence 39, Appl
17	379.5	37.3	1413	1	US-08-707-237A-45	Sequence 45, Appl
18	379.5	37.3	1464	3	US-08-477-509B-74	Sequence 74, Appl
19	379.5	37.3	1464	3	US-08-482-085B-74	Sequence 74, Appl
20	376.5	37.0	281	1	US-08-397-633A-75	Sequence 75, Appl
21	376.5	37.0	287	1	US-08-397-633A-76	Sequence 76, Appl
22	376.5	37.0	877	1	US-08-397-633A-54	Sequence 54, Appl
23	371.5	36.5	2257	1	US-08-175-155-47	Sequence 47, Appl
24	371.5	36.5	2257	1	US-08-477-509B-82	Sequence 82, Appl
25	371.5	36.5	2257	2	US-08-707-237A-53	Sequence 53, Appl
26	371.5	36.5	2257	2	US-08-482-085B-82	Sequence 82, Appl
27	368.5	36.2	378	3	US-08-707-237A-104	Sequence 104, App
28	368.5	36.2	378	3	US-08-642-246-26	Sequence 26, Appl
29	368.5	36.2	378	5	PCT-US96-06229-26	Sequence 26, Appl
30	368.5	36.2	1002	2	US-08-707-237A-103	Sequence 103, App
31	368.5	36.2	1002	1	US-08-642-246-25	Sequence 25, Appl
32	368.5	36.2	1002	5	PCT-US96-06229-25	Sequence 25, Appl
33	366.5	36.0	2055	1	US-08-175-155-46	Sequence 46, Appl
34	366.5	36.0	2055	1	US-08-477-509B-81	Sequence 81, Appl
35	366.5	36.0	2055	2	US-08-707-237A-52	Sequence 81, Appl
36	366.5	36.0	2055	2	US-08-482-085B-81	Sequence 81, Appl
37	363.5	35.7	884	1	US-08-397-633A-68	Sequence 68, Appl
38	363.5	35.7	884	2	US-08-435-641-15	Sequence 15, Appl
39	363.5	35.7	884	2	US-08-707-237A-96	Sequence 96, Appl
40	363.5	35.7	884	3	US-08-642-246-15	Sequence 15, Appl
41	363.5	35.7	884	5	PCT-US96-06229-15	Sequence 15, Appl
42	362	35.6	486	1	US-08-397-633A-77	Sequence 77, Appl
43	361.5	35.5	262	1	US-08-397-633A-73	Sequence 73, Appl
44	358.5	35.3	479	1	US-08-397-633A-78	Sequence 78, Appl
45	357.5	35.2	208	1	US-08-212-237-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-08-911-364-2
; Sequence 2, Application US/08911364
; Patent No. 5969106
; GENERAL INFORMATION:
; APPLICANT: ROTHSTEIN, Aser
; APPLICANT: KEELY, Fred W.
; APPLICANT: ROTHSTEIN, Steven J.
; TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
; TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,364
; FILING DATE: 07-AUG-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,552
; FILING DATE: 07-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 041082/0104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-911-364-2

Query Match 100.0%; Score 1017; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 3.9e-77;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   61 AAAKAAQFGLVPGVGVPGVGVPGVGVGPGVGVPGVGVPGVGVLAPGVGVPGVGVPGVGVPGVGVPAIGPPE 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   438 AAAKAAQFGLVPGVGVPGVGVPGVGVGPGVGVPGVGVLAPGVGVLAPGVGVGPGVGVPGVGVPAIGIGP-- 495

Qy   121 AQAAAAAKAAKYGVGTPAAAAAKAAKAAKAAQFGL---VPGVGVPGVGVPGVGVAPGV-G 175
          ||||:| | ||| .|| || :|| .:||:|| |:|:||:| |
Db   496 GGVAAAAKSA------AKVLAAKAQLRAAA-GLGAGIPGLGV--GVGV-PGLGVGAGVTPG 544

Qy   177 LAPGVGVAPGVGVAPGVGVAPAIGP 201
          | .||||| || | .:||
Db   545 LGVGAGV-PGPGAGAGEGVRRSLSP 568

RESULT 3
US-08-678-039A-40
; Sequence 40, Application US/08678039A
; Patent No. 5858662
;; GENERAL INFORMATION:
;   APPLICANT: Keating, Mark T.
;   APPLICANT: Morris, Colleen A.
;   TITLE OF INVENTION: Diagnosis of Williams Syndrome and
;   TITLE OF INVENTION: Williams Syndrome Cognitive Profile by Analysis of the
;   TITLE OF INVENTION: Presence or Absence of a LIM-Kinase Gene
;   NUMBER OF SEQUENCES: 42
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
;   STREET: 555 Thirteenth Street, N.W., Suite 701 East
;   STREET: Tower
;   CITY: Washington
;   STATE: DC
;   COUNTRY: U.S.A.
;   ZIP: 20004
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/678,039A
;   FILING DATE: 10-JUL-1996
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Saxe, Stephen A.
;   REGISTRATION NUMBER: 38,609
;   REFERENCE/DOCKET NUMBER: 2323-120A
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-624-1589
;   TELEFAX: 202-783-6031
;   INFORMATION FOR SEQ ID NO: 40:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 792 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-678-039A-40

Query Match   65.1%; Score 662.5; DB 2; Length 792;
Best Local Similarity   65.0%; Pred. No. 2e-47;
Matches 156; Conservative   7; Mismatches   53; Indels   24; Gaps   10;

Qy   1 FPGFGVGVGIPGVAGYPGVGVGGVPGVGGVPGVGVGPGVGISPEAQAAAAAKAAKY-----------------49
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   404 FPGFGVGVGIPGVAGYPSVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGAAGAGVLAGL 463

Qy   50 ----------GVGTPAAAAAKAAKAAQP------GLVPGVGVAPGVGVAPG 85
                      ||||||||||||||| ||||||||||||||||
Db   464 VPGPQAAVPGTPGTPGGVPGVGGVGTPAAAAAKAAKAAQFALILNALGLVPGVPGVGVAPGVGVAPG 523

Qy   86 VGVAPGVGLAPGVGVAPGVGVAPGVGVGVGVAPGVGVGVAPGVGVGP--GGVAAAAKAAA 145
          ||||||||||||||||||||||||||||||||| || | :||
Db   524 VGVAPGVGLAPGVGVAPGVGVAPGVGVGVGVAPGVGVGVAPGVGVGP--GGVAAAAKSA------AKVAAAKAQ 574

RESULT 2
US-08-911-364-1
; Sequence 1, Application US/08911364
; Patent No. 5969106
;; GENERAL INFORMATION:
;   APPLICANT: ROTHSTEIN, Aser
;   APPLICANT: KEELY, Fred W.
;   APPLICANT: ROTHSTEIN, Steven J.
;   TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
;   TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FOLEY & LARDNER
;   STREET: 3000 K Street, N.W.
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20007-5109
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/911,364
;   FILING DATE: 07-AUG-1997
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/023,552
;   FILING DATE: 07-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Bent, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 041082/0104
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 672-5300
;   TELEFAX: (202) 672-5399
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 731 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-911-364-1

Query Match   68.8%; Score 700; DB 2; Length 731;
Best Local Similarity   76.1%; Pred. No. 1.5e-50;
Matches 156; Conservative   7; Mismatches   24; Indels   18; Gaps   8;

Qy   1 FPGFGVGVGIPGVAGYPGVGVGGVPGVGGVPGVGVGPGVGISPEAQAAAAAKAAKYGVGTPAAAAAK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   378 FPGFGVGVGIPGVAGYPSVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGVGTPAAAAAK 437

Qy   1 FPGFGVGVGGIPGVAGYPGVGVGGVPGVGVPGVGVGPGVGISPEAQAAAAAKAAKYGVGTPAAAAAK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 FPGFGVGVGAGVPGVGVPGVGVPGVGVGPGVGISPEAQAAAAAKAAKAAKYGVGTPAAAAAK 60

Qy   61 AAAKAAQFGLVPGVGVPGVGVAPGVGVAPGVGVGLAPGVGVAPGVGVAPGVGVAPGVGVAPAIGPPE 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPAIGPPE 120

Qy   121 AQAAAAAKAAKYGVGTPAAAAAKAAKAAKAAQFGLVPGVGVPGVGVAPGVGVAPGVGVGLAPG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AQAAAAAKAAKYGVGTPAAAAAKAAKAAKAAQFGLVPGVGVPGVGVAPGVGVAPGVGVGLAPG 180

Qy   181 VGVAPGVGVAPAIGP 201
          |||||||||||||||
Db   181 VGVAPGVGVAPAIGP 201

```
Qy  146 AKAAQFGLI---VPGVGVAPGVGVAPGVGVAPGV-GLAPGVGVAPGVGVAPGVGVAPAIGP 201
         :|| ||  .:||:|| ||||  |||:|| || :|| |
Db  575 LRAAA-GLGAGIPGLGV--GVGV-PGLGVGAGVPGLGVGAGV-PGFGAGADEGVRRSLSP 629
```

RESULT 4
US-08-212-237-5
; Sequence 5, Application US/08212237
; Patent No. 5606019
; GENERAL INFORMATION:
;   APPLICANT: Cappello, Joseph
;   TITLE OF INVENTION: Synthetic Proteins As Implantables
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: Four Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: U.S.A.
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/212,237
;   FILING DATE: 11-MAR-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Rowland, Bertram I
;   REGISTRATION NUMBER: 20,015
;   REFERENCE/DOCKET NUMBER: A-58847/BIR
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-781-1989
;   TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 988 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-212-237-5

Query Match   39.4%; Score 401; DB 1; Length 988;
Best Local Similarity   56.5%; Pred. No. 7.4e-26;
Matches 122; Conservative 11; Mismatches 49; Indels 34; Gaps 21;

```
Qy   3 GFGVGVGGIPGV----AGVPGVG----GVPGVG:ISPEAQAAAAKAAKYG 50
        | | |||||   :|:    |||||| | |
Db  69 GSGAGAGSVPGVGVPGVGVPGVGVPGVGVGVPGVGV-PGVGGAGSGAGAG 127

Qy  51 VGTPAAAAAKA-----AAKAAQFGLVPGVGVAPGVGVAPGVGLAPGVGVAPGVG 105
        | |:|  .|  |||||:|||| |||| ||||: |||| ||||
Db 128 SGAGAGSGAGSGAGSGAGSGAGSVPGVGV-PGVGV-PGVGV-PGVG 182

Qy 106 VAPGVGVAPAIGPPEAQAAAAKAAKYGVGTPAAAAAKA-RAKAAQFGLVPGVGVAPGVG 164
        | ||||| :|   |:|  :  |||| | | :|| :||| || | ||||
Db 183 V-PGVGV-PGVGGAGSGAGSGAGSGGA-GAGSGAGSGAGAGSGAGSVPGVGV-PGVG 238

Qy 165 VAPGVGVAPGVGLAPGVGVAPGVGVAPGVGTPAIG 200
        || ||||||| | :||||| | :|
Db 239 V-PGVGV-PGVGV-PGVGV-PGVGV-PGVG 268
```

RESULT 5
PCT-US95-02772-5
; Sequence 5, Application PCT/US9502772
; GENERAL INFORMATION:
;   APPLICANT: Protein Polymer Technologies, Inc.
;   TITLE OF INVENTION: Synthetic Proteins As Implantables

;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: Four Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: U.S.A.
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/02772
;   FILING DATE:
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Rowland, Bertram I
;   REGISTRATION NUMBER: 20,015
;   REFERENCE/DOCKET NUMBER: FP-58847-1-PC/BIR
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-781-1989
;   TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 988 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
PCT-US95-02772-5

Query Match   39.4%; Score 401; DB 5; Length 988;
Best Local Similarity   56.5%; Pred. No. 7.4e-26;
Matches 122; Conservative 11; Mismatches 49; Indels 34; Gaps 21;

```
Qy   3 GFGVGVGGIPGV----AGVPGVG----GVPGVGISPEAQAAAAKAAKYG 50
        | | |||||   :|:    |||||| | |
Db  69 GSGAGAGSVPGVGVPGVGVPGVGVPGVGVGVPGVGV-PGVGGAGSGAGAG 127

Qy  51 VGTPAAAAAKA-----AAKAAQFGLVPGVGVAPGVGVAPGVGLAPGVGVAPGVG 105
        | |:|  .|  |||||:|||| |||| ||||: |||| ||||
Db 128 SGAGAGSGAGSGAGSGAGSGAGSVPGVGV-PGVGV-PGVGV-PGVG 182

Qy 106 VAPGVGVAPAIGPPEAQAAAAKAAKYGVGTPAAAAAKA-RAKAAQFGLVPGVGVAPGVG 164
        | ||||| :|   |:|  :  |||| | | :|| :||| || | ||||
Db 183 V-PGVGV-PGVGGAGSGAGSGAGSGGA-GAGSGAGSGAGAGSGAGSVPGVGV-PGVG 238

Qy 165 VAPGVGVAPGVGLAPGVGVAPGVGVAPGVGTPAIG 200
        || ||||||| | :||||| | :|
Db 239 V-PGVGV-PGVGV-PGVGV-PGVGV-PGVG 268
```

RESULT 6
US-08-212-237-7
; Sequence 7, Application US/08212237
; Patent No. 5606019
; GENERAL INFORMATION:
;   APPLICANT: Cappello, Joseph
;   TITLE OF INVENTION: Synthetic Proteins As Implantables
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: Four Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: U.S.A.
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,237
FILING DATE: 11-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A-58847/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 972 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-212-237-7

Query Match   39.1%;   Score 397.5; DB 1; Length 972;
Best Local Similarity 50.0%;   Pred. No. 1.4e-25;
Matches 126; Conservative 8; Mismatches 53; Indels 65; Gaps 21;

Qy   2 PGFGV-GVG----GIPGVAGVPGVG----GVPGVG----AAKAAQPGLVPGVGVAPGVGVAPGVGVLAPG 97
              | |||||| || |:|| |||||||| | ||| ||||| :|||| |||| |||:|| ||
Db   17 PGVPGVPGVPGVSPGVGV-GVPGVPGVPGVPGVPGVPGVGGAGAGSSGAGAGSSGAG 75

Qy   49 YGVGTPAAAAKA-------------AAKAAQPGLVPGVGVAPGVGVAPGVGVLAPG 97
               | |||| | | :|||||| |||| |||| |||:|| |
Db   76 AGSGAGAGSSGAGAGSSGAGAGSVPGVGV-PGVGV-PGVGV-PG 131

Qy   98 VGV----APGVGV----APAIGPPEAQAAAAKAAKYGVGTPAAAAKA- 144
           ||| |||| | :| : :| |
Db   132 VSVPGVGVPGVGVGVPGVPGVPGVPGVPGVGGAGAGSSGAGAGSSGAGAG 191

Qy   145 ------------------AAKAAQPGLVPGVGVAPGVGVAEGVGVAPGVGVLAPGVGVAPGVPGVG 188
                               | | ||||| |||| ||||: ||||| |||| |||
Db   192 SGAGAGSGAGAGSGAGAGSSGAGAGSVPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVG 246

Qy   189 VAPGVGVAPAIG 200
           || | ||||| |:|
Db   247 V-PGVGV-PGVG 256

RESULT 7
PCT-US95-02772-7
; Sequence 7, Application PCT/US9502772
; GENERAL INFORMATION:
APPLICANT: Protein Polymer Technologies, Inc.
TITLE OF INVENTION: Synthetic Proteins As Implantables
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: U.S.A.
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02772
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I

REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: FP-58847-1-PC/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 972 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-02772-7

Query Match   39.1%;   Score 397.5; DB 5; Length 972;
Best Local Similarity 50.0%;   Pred. No. 1.4e-25;
Matches 126; Conservative 8; Mismatches 53; Indels 65; Gaps 21;

Qy   2 PGFGV-GVG----GIPGVAGVPGVG----GVPGVG----AAKAAQPGVGISPEAQAAAAKAAK 48
              | |||||| || |:|| |||||||| | |||||
Db   17 PGVPGVPGVPGVSPGVGV-GVPGVPGVPGVPGVPGVPGVGGAGAGSSGAGAGSSGAG 75

Qy   49 YGVGTPAAAAKA--------------AAKAAQPGLVPGVGVAPGVGVAPGVGVLAPG 97
               | |||| | | :|||||| |||| |||| |||:|| |
Db   76 AGSGAGAGSSGAGAGSSGAGAGSVPGVGV-PGVGV-PGVGV-PG 131

Qy   98 VGV----APGVGV----APAIGPPEAQAAAAKAAKYGVGTPAAAAKA- 144
           ||| |||| | :| : :| |
Db   132 VSVPGVGVPGVGVGVPGVPGVPGVPGVPGVGGAGAGSSGAGAGSSGAGAG 191

Qy   145 -------------------AAKAAQPGLVPGVGVAPGVGVAPGVGVLAPGVGVAPGV-PGVG-PGVG 188
                               | | |||| |||| ||||| |||| |||:|| |||| |
Db   192 SGAGAGSGAGAGSGAGAGSSGAGAGSVPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVG 246

Qy   189 VAPGVGVAPAIG 200
           || | ||||| |:|
Db   247 V-PGVGV-PGVG 256

RESULT 8
US-08-212-237-6
; Sequence 6, Application US/08212237
; Patent No. 5606019
; GENERAL INFORMATION:
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Synthetic Proteins As Implantables
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: U.S.A.
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,237
FILING DATE: 11-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A-58847/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

```
;      LENGTH: 1056 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-212-237-6

Query Match    38.5%;  Score 392;  DB 1;  Length 1056;
Best Local Similarity  52.8%;  Pred. No. 4.3e-25;
Matches 121;  Conservative  7;  Mismatches  57;  Indels  44;  Gaps  18;

Qy    2  PGFGVGVGGIPGVAGVPGVG----GVPGVGISPEAQAAAAAKAAKYGVGT  53
         | ||| | :||| ||||||      :|||||  :  | | |  ||
Db    2  PGVGVPGVGVPGV-GVPGVGVPGVGVPGVGGAGSGAGSGAGSGA  60

Qy   54  PAAAAAKA--------AAKAAQFGLVPGVGVAPGVGVAPGVGLAPGVGVAP 102
             : :| | :| ||   | | | |
Db   61  GAGSGAGSGAGSGAGSGAGSVPGVGV-PGVGV-PGVGV-PGVGV-P 115

Qy  103  GVGVAPGVGVAPAIGPPEAQAAAAAKAAKYGVGTPAAAAAKA----------AAKAAQF 151
         | ||||                                        | :
Db  116  GVGV-PGVGYPGVGGAGSGAGSGAGSGAGSGAGSGAGSGAGSGAGA  174

Qy  152  GLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPAIG 200
         | |||| | :|  :|
Db  175  GSVPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVG 216

RESULT 9
PCT-US95-02772-6
;  Sequence 6, Application PCT/US9502772
;  GENERAL INFORMATION:
;     APPLICANT: Protein Polymer Technologies, Inc.
;     TITLE OF INVENTION: Synthetic Proteins As Implantables
;     NUMBER OF SEQUENCES: 9
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;        STREET: Four Embarcadero Center, Suite 3400
;        CITY: San Francisco
;        STATE: CA
;        COUNTRY: U.S.A.
;        ZIP: 94111-4187
;     COMPUTER READABLE FORM:
;        MEDIUM TYPE: Floppy disk
;        COMPUTER: IBM PC compatible
;        OPERATING SYSTEM: PC-DOS/MS-DOS
;        SOFTWARE: PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;        APPLICATION NUMBER: PCT/US95/02772
;        FILING DATE:
;        CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;        NAME: Rowland, Bertram I
;        REGISTRATION NUMBER: 20,015
;        REFERENCE/DOCKET NUMBER: FP-58847-1-PC/PIR
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE: 415-781-1989
;        TELEFAX: 415-398-3249
;  INFORMATION FOR SEQ ID NO: 6:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH: 1056 amino acids
;        TYPE: amino acid
;        STRANDEDNESS: single
;        TOPOLOGY: linear
;        MOLECULE TYPE: protein
PCT-US95-02772-6

Query Match    38.5%;  Score 392;  DB 5;  Length 1056;
Best Local Similarity  52.8%;  Pred. No. 4.3e-25;
Matches 121;  Conservative  7;  Mismatches  57;  Indels  44;  Gaps  18;
```

```
Qy    2  PGFGVGVGGIPGVAGVPGVG----GVPGVGISPEAQAAAAAKAAKYGVGT  53
         | ||| | :||| ||||||      :|||||  :  | | |  ||
Db    2  PGVGVPGVGVPGV-GVPGVGVPGVGVPGVGGAGSGAGSGAGSGA  60

Qy   54  PAAAAAKA--------AAKAAQFGLVPGVGVAPGVGVAPGVGLAPGVGVAP 102
             : :| | :| ||   | | | |
Db   61  GAGSGAGSGAGSGAGSGAGSVPGVGV-PGVGV-PGVGV-PGVGV-P 115

Qy  103  GVGVAPGVGVAPAIGPPEAQAAAAAKAAKYGVGTPAAAAAKA----------AAKAAQF 151
Db  116  GVGV-PGVGYPGVGGAGSGAGSGAGSGAGSGAGSGAGSGAGSGAGA  174

Qy  152  GLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPAIG 200
Db  175  GSVPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVG 216

RESULT 10
US-08-212-237-8
;  Sequence 8, Application US/08212237
;  Patent No. 5605019
;  GENERAL INFORMATION:
;     APPLICANT: Cappello, Joseph
;     TITLE OF INVENTION: Synthetic Proteins As Implantables
;     NUMBER OF SEQUENCES: 9
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;        STREET: Four Embarcadero Center, Suite 3400
;        CITY: San Francisco
;        STATE: CA
;        COUNTRY: U.S.A.
;        ZIP: 94111-4187
;     COMPUTER READABLE FORM:
;        MEDIUM TYPE: Floppy disk
;        COMPUTER: IBM PC compatible
;        OPERATING SYSTEM: PC-DOS/MS-DOS
;        SOFTWARE: PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;        APPLICATION NUMBER: US/08/212,237
;        FILING DATE: 11-MAR-1994
;        CLASSIFICATION: 435
;     ATTORNEY/AGENT INFORMATION:
;        NAME: Rowland, Bertram I
;        REGISTRATION NUMBER: 20,015
;        REFERENCE/DOCKET NUMBER: A-58847/3IR
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE: 415-781-1989
;        TELEFAX: 415-398-3249
;  INFORMATION FOR SEQ ID NO: 8:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH: 1024 amino acids
;        TYPE: amino acid
;        STRANDEDNESS: single
;        TOPOLOGY: linear
;        MOLECULE TYPE: peptide
US-08-212-237-8

Query Match    38.5%;  Score 391.5;  DB 1;  Length 1024;
Best Local Similarity  47.0%;  Pred. No. 4.6e-25;
Matches 126;  Conservative  8;  Mismatches  57;  Indels  77;  Gaps  21;

Qy    2  PGFGV-GVG----GIPGVAGVPGVG----GVPGVG----CVPGVGISPEAQAAAAARGAAK  48
         | ||  ||    :|||| ||||||    |||||                        :  |
Db   37  PGVGVPGVGVPGVGY-GVPGVGVPGVGVPGVGVGAGSGAGSGAGSGAG  95

Qy   49  YGVGTPAAAAAKA----------AAKAAQFGLVPGVGV----APGVGV----APGVGV-  88
         : |                   :|||| ||||||    |||||    |||||
Db   96  AGSGAGSGAGSGAGSGAGSGAGSVPGVGVPGVGVPGVGVPGVGVP  155

Qy   89  ---APGVGL----APGVGV-----APGVGV----APAIGPPEAQAAAAAKA  129
         | |  :| |  ||||
Db  156  GVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVGGAGAGSGG  215
```

```
Qy  130  AKYGVGTPAAAAAKA--------------AAKAAQFGLVPGVGVAPGVGVAPGVGVA  172
              |  |:| ||                   | ||||| ||||| |||||
Db  216  AGAGSGAGSGAGSGAGSGAGSGAGSGAGSVPGVGV-PGVGV-  272

Qy  173  PGVGLAPSVGVAPGVGVAPGVGVAPAIG  200
         |||| |  || ||||| ||||| |:|
Db  273  PGVGV-PGVGV-PGVGV-PGVG  296
```

RESULT 11
PCT-US95-02772-8
; Sequence 8, Application PC/TUS9502772
; GENERAL INFORMATION:
; APPLICANT: Protein Polymer Technologies, Inc.
; TITLE OF INVENTION: Synthetic Proteins As Implantables
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02772
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: FP-58847-1-PC/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1024 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US95-02772-8

```
Query Match         38.5%;  Score 391.5;  DB 5;  Length 1024;
Best Local Similarity  47.0%;  Pred. No. 4.6e-25;
Matches  126;  Conservative  8;  Mismatches 57;  Indels 77;  Gaps 21;

Qy    2  PGRGV-GVG----GIPGVAGVPGVG----GVPGVG---  48
                                                        |  |
Db   37  PGVGVPGVGVGVTPGV-GVPGVGVPGVGVPGVGVPGVGVGAGSGAGSGAG  95

Qy   49  YGVGTPAAAAAKA---------------AAKAAQFGLVPGVGV----APGVGV-  88
                                                   ||||   |||||
Db   96  AGSGAGAGSGAGSGAGSGAGSGAGSGAGSVPGVGVGVPGVGVPGVGVGV  155

Qy   89  ---APGVGL----APGVGV----APGVGV---APAIGPPEAQAAAAAAKA  129
            |||||    |||||    ||||   :|:|
Db  156  GVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGAGSGAGSG  215

Qy  130  AKYGVGTPAAAAAKA--------------AAKAAQFGLVPGVGVAPGVGVGVGVA  172
              |  |:| ||                   | ||||| |||||
Db  216  AGAGSGAGSGAGSGAGSGAGSGAGSGAGSVPGVGV-PGVGV-  272

Qy  173  PGVGLAPSVGVAPGVGVAPGVGVAPAIG  200
         |||| |  || ||||| ||||| |:|
Db  273  PGVGV-PGVGV-PGVGV-PGVG  296
```

RESULT 12
US-08-212-237-4
; Sequence 4, Application US/08212237
; Patent No. 5606019
; GENERAL INFORMATION:
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Synthetic Proteins As Implantables
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,237
; FILING DATE: 11-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-58847/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-212-237-4

```
Query Match         38.3%;  Score 390;  DB 1;  Length 832;
Best Local Similarity  57.1%;  Pred. No. 5e-25;
Matches  120;  Conservative 10;  Mismatches 34;  Indels 46;  Gaps 23;

Qy    3  GPGVGVGGIPGV----AGVPGVG----GVPGVGISPEAQAAAAAAKAXYG--G  50
         | ||||| |||||
Db   57  GSGAGAGSVPGVGVPGVGVPGVGVPGVGVPGVGV-PGVGGAGAGSGA--G  113

Qy   51  VCTPAAAAKAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGV  110
                                  |||||| ||||| ||||| |||||
Db  114  AGSGAGAGSGAGA------GSVPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGV  162

Qy  111  GVAPAIGPPEAQAAAAAKAAKYGVGTPAAAAAKAAKAAQFGLVPGVGVGVGVAPGVG  170
         |||
Db  163  GY-PGVG-----GAGAGSGAGSGAGA------GSVPGVGV-PGVGV-PGVG  207

Qy  171  VAPGVGLAPGVGVAPGVGVAPGVGVAPAIG  200
         | |||| ||||| ||||| |:|
Db  208  V-PGVGV-PGVGV-PGVGV-PGVG  232
```

RESULT 13
PCT-US95-02772-4
; Sequence 4, Application PC/TUS952772
; GENERAL INFORMATION:
; APPLICANT: Protein Polymer Technologies, Inc.
; TITLE OF INVENTION: Synthetic Proteins As Implantables
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: U.S.A.
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02772
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: FP-58847-1-PC/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-02772-4

Query Match   38.3%; Score 390; DB 5; Length 832;
Best Local Similarity 57.1%; Pred. No. 5e-25;
Matches 120; Conservative 10; Mismatches 34; Indels 46; Gaps 23;

Qy   3 GFGVGVGGIPGV----AGVPGVG----GVPGVG   50
          | |-|:||| ||||||: | | |
Db   57 GSGAGAGSVPGVPGVGVPGVGVPGVGVPGVGVPGVG   113

Qy   51 VGTPAAAAKAAAAKAAAQFGLVPGVGVAPGVGVAPGVGVAPGV   110
          || |:|| | ||||||| |||| ||||| |||||| |||
Db 114 AGSGAGAGSGAGA-----GSVPGVGV-PGVGV-PGVGV-PGV   162

Qy 111 GVAPAIGPPEAQAAAAKAAKYGVGTPAAAAAKAAAKAAAQFGLVPGVGVAPGVGVAPGVG   170
          ||-|: | |: |-| || | :| | | |||| ||||| ||||
Db 163 GV-PGVG----GAGSGAGSGA--GAGSGAGSGAGA-----GSVPGVGV-PGVGV-PGVG   207

Qy 171 VAPGVGLAPGVGVAPGVGVAPGVGVAPAIG   200
          | |||||: ||||| |||||| ||||: :|
Db 208 V-PGVGV-PGVGV-PGVGV-PGVGV-PGVG   232

RESULT 14
US-08-212-237-3
; Sequence 3, Application US/08212237
; Patent No. 5606019
; GENERAL INFORMATION:
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Synthetic Proteins As Implantables
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/212,237
FILING DATE: 11-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A-58847/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 936 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-212-237-3

Query Match   37.6%; Score 382; DB 1; Length 936;
Best Local Similarity 50.2%; Pred. No. 2.5e-24;
Matches 118; Conservative 41; Mismatches 8; Indels 68; Gaps 21;

Qy   3 GFGVGVGGIPGVAPGVGVG----GVPGVG----GVPGVGISPEAQAAAAAKAAKYGVGTP   54
          | |-|:|| :||| ||||| ||||| ||||:: |||| |
Db   45 GSGAGAGSVPGV-GVPGVGVPGVGVPGVGVP-----------------GVGVP   89

Qy   55 AAAAKAAKAAAQFGLVPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAP   114
          | | || | | ||||| ||||| |||||| |||||| |||||| ||||| |
Db   90 GVGGA-GAGSGAGSVPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVGV-P   141

Qy 115 AIGPPEAQAAAAKAAK----------------------YGVGTPAAAAKAA   145
          |: | ::: |: |||
Db 142 GVGGAGAGSGAGAGSVPGVGVGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPAIG   200

Qy 146 AKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPAIG   200
          |||||| ||||||| |||||| |||||| |||||| | |||:
Db 201 GSGAGAGSVPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVG   248

RESULT 15
PCT-US95-02772-3
; Sequence 3, Application PC/TUS9502772
; GENERAL INFORMATION:
; APPLICANT: Protein Polymer Technologies, Inc.
; TITLE OF INVENTION: Synthetic Proteins As Implantables
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02772
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: FP-58847-1-PC/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

```
;     LENGTH: 936 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
PCT-US95-02772-3


Query Match       37.6%;   Score 382; DB 5;  Length 936;
Best Local Similarity   50.2%;   Pred. No. 2.5e-24;
Matches 118;   Conservative  8;   Mismatches  41;   Indels  68;   Gaps  21;

Qy     3 GFGVGVGGIPGVAGVPGVG----GVPGVG---GVPGVGISPEAQGAAAAAKRAAKYGVGTP 54
           | | || | :||| || ||       || |||| :
Db    45 GSGAGAGSVPGV-GVPGVGVPGVGVPGVGVPGVGVP-------------GVGVP 89

Qy    55 AAAAAKAAAKAAQFGLVPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAP 114
            | | ||||| ||||| ||||| |||| ||||| |||||| |
Db    90 GVGGA-GAGSGAGSGVPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-P 141

Qy   115 AIGPPEAQAAAAKAAK---------------------YGVGTPAAAAAKAA 145
            :| :| : : ||                          ||||| | :
Db   142 GVGGAGSGAGSGAGSGVPGVGVPGVCVPGVGVPGVGVPGVGVPGVGGA-GA 200

Qy   146 AKAAQFGLVPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAPAIG 200
            | | |||||| ||||| |||||| ||||| |||||| |||| |. :|
Db   201 GSGAGAGSVPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVG 248


Search completed: April 24, 2001, 16:36:22
Job time: 302 sec
```

```
            GenCore version 4.5
       Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   April 24, 2001, 16:33:30 ; Search time 125.5 Seconds
                                      (without alignments)
                                      682.701 Million cell updates/sec

Title:          us-09-340-736-1
Perfect score:  3785
Sequence:       1 GGVPGAIPGGVPGGVFYPGA.......LSPIFPGGACLGKACGKKK 731

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

checked:        374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters:   374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :
          SPTREMBL_15:*
       1: sp_archea:*
       2: sp_bacteria:*
       3: sp_fungi:*
       4: sp_human:*
       5: sp_invertebrate:*
       6: sp_mammal:*
       7: sp_mhc:*
       8: sp_organelle:*
       9: sp_phage:*
      10: sp_plant:*
      11: sp_rodent:*
      12: sp_unclassified:*
      13: sp_vertebrate:*
      14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	ID	Description
1	3785	100.0	757	4	Q14234	Q14234	homo sapien
2	3588.5	94.8	724	4	Q14233	Q14233	homo sapien
3	3362	88.8	687	4	Q14235	Q14235	homo sapien
4	3066.5	81.0	635	4	Q15336	Q15336	homo sapien
5	2870	75.8	602	4	Q15337	Q15337	homo sapien
6	2510.5	66.3	707	6	Q28098	Q28098	bos taurus
7	2413.5	63.8	679	6	Q28097	Q28097	bos taurus
8	2262	59.8	666	6	Q28096	Q28096	bos taurus
9	2214	58.5	650	6	Q28099	Q28099	bos taurus
10	1258.5	33.2	258	4	Q9UMF5	Q9UMF5	homo sapien
11	759.5	20.1	988	5	Q17434	Q17434	nephila cla
12	756	20.0	1729	5	Q9U617	Q9U617	drosophila
13	744.5	19.7	1884	5	Q9NHW2	Q9NHW2	nephila mad
14	723	19.1	2249	5	Q9NHW4	Q9NHW4	nephila cla
15	698	18.4	907	5	Q44359	Q44359	nephila cla
16	675.5	17.8	1758	6	Q19098	Q19098	caenorhabdi
17	675.5	17.8	1759	6	Q19099	Q19099	caenorhabdi
18	671	17.7	871	5	Q44358	Q44358	nephila cla
19	666	17.6	1802	5	Q17163	Q17163	brugia mala
20	635	16.8	2639	5	O76786	O76786	antheraea p
21	626	16.5	1489	2	O53559	O53559	mycobacteri
22	612.5	16.2	922	5	O44367	O44367	mytilus edu
23	611	15.8	749	5	O53844	O53844	mycobacteri
24	599	15.7	626	5	Q9NHW1	Q9NHW1	nephila mad
25	595.5	15.7	1306	2	O53775	O53775	mycobacteri
26	594.5	15.7	882	2	O53845	O53845	mycobacteri
27	591	15.6	1660	2	O53215	O53215	mycobacteri
28	583	15.4	1329	2	O68810	O68810	mycobacteri
29	580.5	15.3	738	2	O02402	O02402	pinctada fu
30	580	15.3	1461	4	O76045	O76045	homo sapien
31	578	15.3	1447	13	Q91B91	Q91B91	xenopus lae
32	576	15.2	617	5	O46172	O46172	nephila cla
33	573	15.1	767	2	O53435	O53435	mycobacteri
34	571	15.1	1160	4	Q14046	Q14046	homo sapien
35	571	15.1	1487	4	Q14047	Q14047	homo sapien
36	569	15.0	1039	5	Q9VTR6	Q9VTR6	drosophila
37	567	15.0	783	2	O53809	O53809	mycobacteri
38	567	15.0	1418	6	Q28396	Q28396	equus cabal
39	566.5	15.0	1690	4	Q12823	Q12823	homo sapien
40	566.5	15.0	1690	5	Q90M96	Q90M96	homo sapien
41	566.5	15.0	1691	4	Q9Y414	Q9Y414	homo sapien
42	566	15.0	904	5	O76271	O76271	mytilus edu
43	565	14.9	1372	11	Q9RIE8	Q9RIE8	rattus norv
44	565	14.9	1453	11	O63079	O63079	rattus norv
45	564.5	14.9	812	5	Q06452	Q06452	ephydatia m

ALIGNMENTS

```
RESULT   1
Q14234
ID   Q14234        PRELIMINARY;      PRT;   757 AA.
AC   Q14234;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1995 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN.
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87274906; PubMed=3038460;
RA   Indik Z., Yoon K., Morrow S.D., Cicila G., Rosenbloom J.,
RA   Rosenbloom J., Ornstein-Goldstein N.;
RT   "Alternative splicing of human elastin mRNA indicated by sequence
RT   analysis of cloned genomic and complementary DNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87289668; PubMed=3039501;
RA   Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT   "Structure of the 3' region of the human elastin gene: great abundance
RT   of Alu repetitive sequences and few coding sequences.";
RL   Connect. Tissue Res. 16:197-211(1987).
DR   EMBL; M17282; AAC98395.1; -.
DR   EMBL; M15983; AAC98395.1; JOINED.
DR   EMBL; M17265; AAC98395.1; JOINED.
DR   EMBL; M17266; AAC98395.1; JOINED.
DR   EMBL; M17267; AAC98395.1; JOINED.
DR   EMBL; M17268; AAC98395.1; JOINED.
DR   EMBL; M17270; AAC98395.1; JOINED.
DR   EMBL; M17271; AAC98395.1; JOINED.
DR   EMBL; M17272; AAC98395.1; JOINED.
DR   EMBL; M17273; AAC98395.1; JOINED.
DR   EMBL; M17274; AAC98395.1; JOINED.
DR   EMBL; M17275; AAC98395.1; JOINED.
DR   EMBL; M17276; AAC98395.1; JOINED.
```

```
DR   EMBL; M17277; AAC98395.1; JOINED.
DR   EMBL; M17278; AAC98395.1; JOINED.
DR   EMBL; M17279; AAC98395.1; JOINED.
DR   EMBL; M17280; AAC98395.1; JOINED.
DR   EMBL; M17281; AAC98395.1; JOINED.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
SQ   SEQUENCE   757 AA;  66136 MW;  23B7FE5B8AF95CA8 CRC64;
```

```
Query Match             100.0%;  Score 3785;  DB 4;  Length 757;
Best Local Similarity   100.0%;  Pred. No. 1.6e-174;
Matches 731;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy    1 GGVPGGAIPGGVPGGVYPGAGLGALGGALGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27 GGVPGGAIPGGVPGGVYPGAGLGALGGALGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT 86

Qy   61 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL 146

Qy  121 PGVYPGGVLPGGARPPGVGVLPGVPTGAGVKPAPGVGAFAGIPGVGPFGGPGPGVPLGY 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  147 PGVYPGGVLPGGARPPGVGVLPGVPTGAGVKPAPGVGAFAGIPGVGPFGGPGPGVPLGY 206

Qy  181 PIKAPKLPGGYGLPYTTGKLPYVGVPGYGPGGVAGGKAGYPTGTCVGPQRAAAAAKAAAKF 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  207 PIKAPKLPGGYGLPYTTGKLPYVGYGPGGVAGGKAGYPTGTGVGPQRAAAAAAKAAAKF 266

Qy  241 GAGAAGVLPGGVAPGVPGCAIPGIGGIAGVGTPAAAAAAAAAAKAAKYGAAAGLVPGG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  267 GAGAAGVLPGGVAPGVPGCAIPGIGGIAGVGTPAAAAAAAAAAKAAKYGAAAGLVPGG 326

Qy  301 PGFGPGVVGPGAGVPGVGVPGAGIPPVVPGAGIPPGAAVPGVYSPAAAKAAKAAKYGAR 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  327 FGFGPGVVGPGAGVPGVGVPGAGIPGAAVPGVYSPBAAAKAAAKAAKYGAR 386

Qy  361 PGVGVGIPPTVGAGGPGPGVGVGGIPGVAGVPSVGGVPGVGVPGVGIPEAQAAAA 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  387 PGVGVGGIPPTVGAGGPGPGVGVGGIPGVAGVPSVGGVPGVGVGIPEAQAAAA 446

Qy  421 AKAAKYGVSTPAAAAAKAAAKAAQFGVLGVGGVCVAPGVGLAPGVGVAPG 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  447 AKAAKYGVGTPAAAAAKAAAKAAQFGVLGVGGVCVAPGVGLAPGVGVAPG 506

Qy  481 VGVAPGCVAPGIGPGGVAAAAKSAAKYAAKAQLPRAAGLGAGIPGLGVGVPGCLGVGA 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  507 VGVAPGCVAPGIGPGGVAAAAKSAAKYAAKAQLPRAAGLGAGIPGLGVGVPGLGVGA 566

Qy  541 GVPGLGVGAGVPGFGFGAGADEGVRRSLSPELRSGDPSSSQHLPSTPSSPRVPGALAAAKAA 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  567 GVPGLGVGAGVPGFGFGAGADEGVRRSLSPELRSGDPSSSQHLPSTPSSPRVPGALAAAKAA 626

Qy  601 KYGAAVPGVLGGLGALGGVGIPGGVVPGAGPAAAAAAKTAAAAKAQPGLVGAAGLGGLGVGVG 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  627 KYGAAVPGVLGGLGALGGVGIPGGVVPGAGPAAAAAAKTAAAAKAQPGLVGAAGLGGLGVGVG 686

Qy  661 GLGVPGVGLGGIPPAAAAAAKYGAAGLGGVLPGGAGQFPLGGVAARPGLSPIFPGGA 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  687 GLGVPGVGLGGIPPAAAAAAKYGAAGLGGVLGGAGQFPLGGVAARPGLSPIFPGGA 746

Qy  721 CLGKACGRKRK 731
         |||||||||||
Db  747 CLGKACGRKRK 757
```

```
RESULT  2
Q14233
ID   Q14233          PRELIMINARY;          PRT;   724 AA.
AC   Q14233; Q14238;
```

```
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN.
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87274906; PubMed=303460;
RA   Indik Z., Yoon K., Morrow S.D., Cicila G., Rosenbloom J.,
RA   Rosenbloom J., Ornstein-Goldstein N.;
RT   "Structure of the 3' region of the human elastin gene: great abundance
RT   of Alu repetitive sequences and few coding sequences.";
RL   Connect. Tissue Res. 16:197-211(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87289668; PubMed=3039501;
RA   Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT   "Alternative splicing of human elastin mRNA indicated by sequence
RT   analysis of cloned genomic and complementary DNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
RN   [3]
RP   SEQUENCE OF 164-724 FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE=88156138; PubMed=2831431;
RA   Fazio M.J., Olsen D.R., Kuivanieni H., Chu M.L., Davidson J.M.,
RA   Rosenbloom J., Uitto J.;
RT   "Isolation and characterization of human elastin cDNAs, and age-
RT   associated variation in elastin gene expression in cultured skin
RT   fibroblasts.";
RL   Lab. Invest. 58:270-277(1988).
DR   EMBL; M17282; AAC98394.1; -.
DR   EMBL; M16983; AAC98394.1; JOINED.
DR   EMBL; M17265; AAC98394.1; JOINED.
DR   EMBL; M17266; AAC98394.1; JOINED.
DR   EMBL; M17267; AAC98394.1; JOINED.
DR   EMBL; M17269; AAC98394.1; JOINED.
DR   EMBL; M17270; AAC98394.1; JOINED.
DR   EMBL; M17271; AAC98394.1; JOINED.
DR   EMBL; M17272; AAC98394.1; JOINED.
DR   EMBL; M17273; AAC98394.1; JOINED.
DR   EMBL; M17275; AAC98394.1; JOINED.
DR   EMBL; M17276; AAC98394.1; JOINED.
DR   EMBL; M17277; AAC98394.1; JOINED.
DR   EMBL; M17278; AAC98394.1; JOINED.
DR   EMBL; M17279; AAC98394.1; JOINED.
DR   EMBL; M17280; AAC98394.1; JOINED.
DR   EMBL; M17281; AAC98394.1; JOINED.
DR   EMBL; M24782; AAA53190.1; -.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
SQ   SEQUENCE   724 AA;  62664 MW;  6155686228EDF3D CRC64;
```

```
Query Match             94.8%;  Score 3588.5;  DB 4;  Length 724;
Best Local Similarity   95.5%;  Pred. No. 4e-165;
Matches 698;  Conservative   0;  Mismatches   0;  Indels   33;  Gaps   1;
```

```
Qy    1 GGVPGGAIPGGVPGGVYPGAGLGALGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27 GGVPGGAIPGGVPGGVYPGAGLGALGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT 86

Qy   61 FPGALVPGGVADAAAATKAAKAGAGLAGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL 120
         |||||||||||||   ||||||||||| ||||||||||||||||||||||||||||||
Db   87 FPGALVPGGVADAAAAYKAAKAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGKVPGVGL 146

Qy  121 PGVYPGGVLPGGARPPGVGVLPGVPTGAGVKPKAPGVGGAFAGIPGVGPFGGPGPGVPLGY 180
         |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  121 PGVYPGGVLPGGARPPGVGVLPGVPTGAGVKPAPGVGGAFAGIPGVGPFGGPGPGVPLGY 180
```

```
DR   EMBL; M17282; AAC98393.1; -.
DR   EMBL; M16983; AAC98393.1; JOINED.
DR   EMBL; M17265; AAC98393.1; JOINED.
DR   EMBL; M17266; AAC98393.1; JOINED.
DR   EMBL; M17267; AAC98393.1; JOINED.
DR   EMBL; M17268; AAC98393.1; JOINED.
DR   EMBL; M17271; AAC98393.1; JOINED.
DR   EMBL; M17272; AAC98393.1; JOINED.
DR   EMBL; M17273; AAC98393.1; JOINED.
DR   EMBL; M17275; AAC98393.1; JOINED.
DR   EMBL; M17276; AAC98393.1; JOINED.
DR   EMBL; M17277; AAC98393.1; JOINED.
DR   EMBL; M17278; AAC98393.1; JOINED.
DR   EMBL; M17279; AAC98393.1; JOINED.
DR   EMBL; M17281; AAC98393.1; JOINED.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES_UNKNOWN_1.
DR   PROSITE; PS00453; FKBP_PPIASE_1 UNKNOWN_1.
SQ   SEQUENCE   687 AA;  59579 MW;  9D5AC3C4D9F9E93B CRC64;
```

```
Query Match             88.8%;  Score 3362;  DB 4;  Length 687;
Best Local Similarity   90.3%;  Pred. No. 2.7e-154;
Matches  660;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    3;

Qy         1 GGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        27 GGVPGAIPGGVPGGVFYPGAGLGALGGGALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVT 86

Qy        61 FPGALVPGGVADAAAAYKAAKAAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGVPGVGL 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        87 FPGALVPGGVADAAAAYKAAKAAGAGLGGVPGVGGLGVSAGAVVPQPGAGVKPGVPGVGL 146

Qy       121 PGVYPGGVLPGARFPGVGVLPGVPTGAGVGAAGKAGYPTGTGVGPQPGPGVPELGY 180
             ||||||||||||||||||||||||||||||
Db       147 PGVYPGGVLPGARFPGVGVLPGVPTGAGVGAAFAGIPGVGPGGQPGVPLGY 206

Qy       181 PIKAPKLPGGYGLPYTTGKLPYTGSPGPGVGAGKAGYPTGVGPQAAAAAAKAAAKF 240
             ||||||||||||||||||||||||||||||||||||||||||||||
Db       207 PIKAPKLPGGYGLPYTTGKLPYTGSPGPGVGAGKAGYPTGVGPQAAAAAAAAKAAAKF 266

Qy       241 GAGAAGVLPGVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAKAAKYGAAAGLVPGG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       267 GAGAAGVLPGVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAKAAKYGAAAGLVPGG 326

Qy       301 PGRGPGVVGVGVVPGAGVGVGVYPGVGVPGIPVVPGAGIPGAAVPGVVSPEAAAKAAKYGAR 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       327 PGRGPGVVGVGVVPGAGVGVGVYPGVGVPGIPVVPGAGIPGAAVPGVVSPEAAAKAAKYGAR 386

Qy       361 PGVGVGGIPTYGVGAGGFPGFGVGVGGISPTAGVGVGGVPGVGISPEAQAAAA 420
             ||||||||||||||||||||||||||||||||||||||||||||||
Db       387 PGVGVGGIPTYGVGAGGFPGFGVGVGGISPTAGVGVGGVPGVGISPEAQAAAA 446

Qy       421 AKAAKYGVGTPAAAAAAKAAAKAAAKAAQPGLVPGVGVAPGVGVAPGVGLAPGVGVAPG 480
             ||||||
Db       447 AKAAKY------------GLVPGVGVAPGVGVAPGVGLAPGVGVAPG 487

Qy       481 VGVAPGVGVGPGGVAAAKAAQFGLVPGVGVAAKVAAKAQLRAAGLGAGIPGLGVGVGGVPGLGVGA 540
             ||||||||||||           ||||||||||  ||||   |||||||||||||||||||||||
Db       488 VGVAPGVGVAPGIGPGGVAAAAKSAAKVRAKAQLRAAAGLGAGIPGLGVGVGGVPGLGVGA 547

Qy       541 GVPGLGVGAGVPGFGAGADGEVRKSLSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAA 600
             ||||||||||||||||                                  ||||||||||
Db       548 ------------------------------VPGALAAAKAA 574

Qy       601 KYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAAKAAAKAAQFGLVGAAGLGGLGVG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       575 KYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAAKAAAKAAQFGLVGAAGLGGLGVG 634

Qy       661 GLGVPGVGGLGGIPPAAAAKAAKYGAAGLGGVLGGAGQFPLLGGVAARPGFGLSPIPPGGA 720
             ||||||||||||||||||||||||                     |||||||||||||||
Db       635 GLGVPGVGGLGGIPPAAAAKAAKY------------------------GVAARPGFGLSPIPPGGA 676
```

```
Db       147 PGVTPCGVLPGARFPGVGVLPGVLPGVPTCAGVKPKAPGVGGAFAGIPGVGPGFPGGPQPGVPLGY 206

Qy       181 PIKAPKLPGGYGLPYTTGKLPYTGSPGPGVGAGKAGYPTGVGPQAAAAAAKAAAKF 240
             |||||||||||||||||||||||||||||||||||||||||||||||
Db       207 PIKAPKLPGGYGLPYTTGKLPYTGSPGPGVGAGKAGYPTGVGPQAAAAAAAAKAAAKF 266

Qy       241 GAGAAGVLPGVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAKAAKYGAAAGLVPGG 300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       267 GAGAAGVLPGVGGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAKAAKYGAAAGLVPGG 326

Qy       301 PGFGPGVVGVVPGAGVGVGVYPGVGVPGIPVVPGAAVPGVVSPEAAAKAAAKAAKYGAR 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       327 PGPGPGVVVPGVGVGVSVPGAGVGVGIPVVPGAGVPGVVSPEAAAKAAAKAAKYGAR 386

Qy       361 PGVGVGGIPTYGVGAGVPSVGGVGVGVPGVGVPGVGVPGVPGVGISPEAQAAAA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       387 PGVGVGGIPTYGVGAGVPSVGGVGVGVPGVGVPGVGVPGVPGVGISPEAQAAAA 446

Qy       421 AKAAKYGVGTPAAAAAAKAAAKAAAKAAQPGLVPGVGVAPGVGVAPGVGLAPGVGVAPG 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       447 AKAAKYGVGTPAAAAAAKAAAKAAAKAAQPGLVPGVGVAPGVGVAPGVGLAPGVGVAPG 506

Qy       481 VGVAPGVGVAGVPGFPGAVPGVLGGLGALGGVGIPGGVVGVGVPGLGVGA 540
             ||||||||||||||||||||||||||||||||||||||||||
Db       507 VGVAPGVGVAGVPGFPGAGGFPGFGVGVGGLGALGGVGIPGGVVGVGVPGLGVGA 566

Qy       541 GVPGLGVGAGVPGFPGAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAA 600
             |||||||||||||||
Db       567 GVPGLGVGAGVPGFGA------------------------VPGALAAAKAA 593

Qy       601 KYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAAKAAAKAAQPGLVGAAGLGGLGVG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       594 KYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAAKAAAKAAQPGLVGAAGLGGLGVG 653

Qy       661 GLGVPGVGGLGGIPPAAAAKAAKYGAAGLGGVLGGAGQPPLLGGVAARPGFGLSPIPPGGA 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       654 GLGVPGVGGLGGIPPAAAAKAAKYGAAGLGGVLGGAGQPPLLGGVAARPGFGLSPIPPGGA 713

Qy       721 CLGACGRKRK 731
             ||||||||||
Db       714 CLGACGRKRK 724
```

```
RESULT  3
O14235
O14235     PRELIMINARY;    PRT;   687 AA.
Q14235;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN.
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87274906; PubMed=3018460;
RA   Indik Z., Yoon K., Morrow S.D., Cicila G., Rosenbloom J.,
RA   Rosenbloom J., Ornstein-Goldstein N.;
RT   "Structure of the 3' region of the human elastin gene: great abundance
RT   of Alu repetitive sequences and few coding sequences.";
RL   Connect. Tissue Res. 16:197-211(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87289668; PubMed=3039501;
RA   Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT   "Alternative splicing of human elastin mRNA indicated by sequence
RT   analysis of cloned genomic and complementary DNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
```

Qy   721 CLGKACGRKRK 731
          ||||||||||
Db   677 CLGKACGRKRK 687

RESULT 4
O15336
ID   O15336;   PRELIMINARY;   PRT;   635 AA.
AC   O15336;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Li D.Y., Toland A.E., Boak B.B., Atkinson D.L., Ensing G.J.,
RA   Morris C.A., Keating M.T.;
RL   Hum. Mol. Genet. 0:0-0(1997).
DR   EMBL; U93037; AAB65621.1; -.
DR   EMBL; U93034; AAB65621.1; JOINED.
DR   EMBL; U93035; AAB65621.1; JOINED.
DR   EMBL; U93036; AAB65621.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   INTERPRO; IPR001807; -.
DR   INTERPRO; IPR002185; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00569; DOPAMINED4R.
DR   PRINTS; PR00762; CLCHANNEL.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
FT   NON_TER   1   1
FT   NON_TER   635   635
SQ   SEQUENCE   635 AA; 55279 MW; 729S0C3641127B2A4 CRC64;

Query Match   81.0%; Score 3066.5; DB 4; Length 635;
Best Local Similarity   94.3%; Pred. No. 3.7e-140;
Matches 599; Conservative   0; Mismatches   1; Indels   35; Gaps   2;

Qy   3 VPGAIPGGVPGGVFYPGAGLGALGGGALGP2GGKPLKPVPGGLAGAGLGAGLGAFPAVTFP 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 VPGAIPGGVPGGVFYPGAGLGALGGGALGP2GGKPLKPVPGGLAGAGLGAGLGAFPAVTFP 60

Qy   63 GALVPGGVADAAAAYKAAKAGAGLGGVPGVGGIPGIGAGVP2GAAAAAAAKAAAKYGAAGLVPGGPG 122
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GALVPGGVADAAAAYKAAKAGAGLGGVPGVGGIPGIGAGVP2GAAAAAAAKAAAKYGAAGLVPGGPG 120

Qy   123 VIPGGVLPGARFPGVGVLPGVPTGAGVKPKAPGVGGRAPVGVGPPGGPGVPLGYPI 182
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   121 VIPGGVLPGARFPGVSVLPGVPTGAGVKPAPGVGGRAPVGVGPPGGPGVPLGYPI 180

Qy   183 KAPKLPGGYGLPYTTGKLPYGYGPGVAGAAGKAGYPTGYGVPGVAGAAAKARARKFGA 242
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 KAPKLPGGYGLPYTTGKLPYGYGPGVAGAAGKAGYPTGYGVPGVAGAAAKARARKFGA 240

Qy   243 GAAGVLPGVSGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAAAKYGAAAGLVPGGPG 302
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAAGVLPGVSGAGVPGVPGAIPGIGGIAGVGTPAAAAAAAAAAAKYGAAAGLVPGGPG 300

Qy   303 FGPGVVGVPGAGVPGVPGAIPVVPGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAAKYGARPG 362
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 FGPGVVGVPGAGVPGVPGAIPGIGGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAAKYGARPG 360

Qy   363 VGVGIPTYGVGAGGFPGFPGFGEGVGIPGVAGVPGVGGVPGVGGVPTGVGIPGTVGIPTVGGIPTYVGGIPTYVGGIPSVGVGIPSPEAQAAAAK 422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 VGVGIPTYGVGAGGFPGFPGFGEGVGIPGVAGVPGVGGVPSVGVGIPSVGVGIPSPEAQAAAAK 420

```
Qy   243 GAAGVLPGVGAGVPGVGPGAIPGIGGIAGVGTPAAAAAAAAAKRAAKYGAAAGLVPGGPG 302
             |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAAGVLPGVGCAGVPGVFGVFGAIPGIGGIAGVGTPAAAAAAAAAKRAAKYGAAAGLVPGGPG 300

Qy   303 FGPGVVGVPGAGVPGVGVPGVPVPGAGIPGAAVPGVVSPERAAAKAAKAAKYGARPG 362
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 FGPGVVGVPGAGVPGVGVPGVPVPGAGIPGAAVPGVVSPERAAAKAAKAAKYGARPG 360

Qy   363 VGVGGIPPIGVGAGGPGGPGVGVGGIPGVGVGGVPGVSVGGVPGVGGIPGVGVGGIPEAQAAAAAK 422
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 VGVGGIPPIGVGAGGPGGPGVGVGGIPGVGVGGVPGVSVGGVPGVGGIPGVGVGGIPEAQAAAAAK 420

Qy   423 AAKY------------------GVGTPAAAAAKAAAKAQF------GL 447
             ||||                                     ||
Db   421 AAKYGAAGAVYLGGLVPGAPGAVPGVPGGTGGVPGVGGTPAAAAAKRAAKAAKAQFALLNLLAGL 480

Qy   448 VPGVGVAPGVGVAPGVGVGLAPGVGVGVAPGVGVGVAPGVGVGIGGPGGVAAAKSAAK 507
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 VPGVGVAPGVGVAPGVGVGLAPGVGVAPGVGVGVAPGVGVGIGGPGGVAAAKSAAK 540

Qy   508 VAAKAQLRAAAGLGAGIPGLGVGVPGGLGVGVGLGVGAGVPGFGAGADDGVRRSLS 567
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 VAAKAQLRAAAGLGAGIPGLGVGVPGGLGVGVGLGVGAGVPGFGAGVPGFGFRA----------- 589

Qy   568 PELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKY 602
             ||||||||||||||||||||||||||||||||||
Db   590 --------------------------VPGALAAAKAAKY 602
```

RESULT 6

```
ID   Q28098            PRELIMINARY;     PRT;   707 AA.
AC   Q28098;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992576;
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J., Yoon K.;
RT   "Structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Cicila G., Yoon K., Rosenbloom J.;
RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30498.1; JOINED.
DR   EMBL; M11422; AAA30498.1; JOINED.
DR   EMBL; M19367; AAA30498.1; JOINED.
DR   EMBL; M19366; AAA30498.1; JOINED.
DR   EMBL; M19368; AAA30498.1; JOINED.
DR   EMBL; M19369; AAA30498.1; JOINED.
DR   EMBL; M19370; AAA30498.1; JOINED.
DR   EMBL; M19371; AAA30498.1; JOINED.
DR   EMBL; M22771; AAA30498.1; JOINED.
DR   EMBL; M22772; AAA30498.1; JOINED.
DR   EMBL; M22773; AAA30498.1; JOINED.
DR   EMBL; M22774; AAA30498.1; JOINED.
DR   EMBL; M22775; AAA30498.1; JOINED.
DR   EMBL; M22988; AAA30498.1; JOINED.
DR   EMBL; M23010; AAA30498.1; JOINED.
```

```
DR   INTERPRO; IPR000104; -.
DR   INTERPRO; IPR001459; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00959; MEVGALKINASE.
FT   NON_TER         1      1
SQ   SEQUENCE   707 AA;   60346 MW;  FDFD559BAB34CE33 CRC64;
```

```
Query Match          66.3%;  Score 2510.5;  DB 6;  Length 707;
Best Local Similarity  68.6%;  Pred. No. 1.8e-113;
Matches 532;  Conservative 29;  Mismatches 74;  Indels 141;  Gaps 22;

Qy    28 GALGPGGKPLKPVPGGLAGAGLGAGLGAFPAVTFPGALVPGGVADAAAAYK-AAKAGA-- 84
            | |||| ||   ||  | | | ||||||||||||||||||||||||||||  |||||
Db     1 GGLGPGGVGPAAKPGVGGLVGPGLGAGLGALPG-AFPGALVPGGPAGAAAAYKAAAKAGAGG 59

Qy    85 -GLGGVPGVGGLGVSAGAVVPQ----PGAGVXPGKVPGGVGLPGVYPGGVLP--GARPGV 137
            |||||||| ||| |||||:||       ||||||||||||||:|||| |  |||||||
Db    60 LGVGGIGGVGGLGVSTGAVVPQLGAGVGAGVPGKVPGKVPGGVGLPGVLPGVLPGAGARPGI 119

Qy   138 GVLPGVPGAGVKPKAGCVGGARPAGIPGVGPPGGPQGPGVPLGVGTPIXAPKLPGGYGLPYTI 197
            |||||||  |||||||::|||||||||:|||||||||||||||||||||||||||:|||||
Db   120 GVLPGVPGAGVKPAPGGGPAPAGIPGVGPPGGXQGVPLGYPIKAPKLPAGYGLPYKT 179

Qy   198 GKLPYCXPGGVAGAAGKAGYPTGYCGVPGPGAAAAAAAAAAAAKFGAGAAGVLP---GVGGAG 255
            |||||||||||||||||||||||||| |||||||||||||||||||||| | ||     |||||
Db   180 GKLPYCXRPGGVVAGSAGKAGYPTGYGVPGPQ-AAAAAKAAAKLGAGGAGVLPGVGVGGAG 238

Qy   256 VPGVPGAIPGIGGIAGVGVTPAAAAAAAAAAAKYGAAAGL---VP--GGPGPG-RGV- 307
            :||  |||||||||||||||   |||||||||||||||  |||       | | | ||
Db   239 IPGAPGAIPGIGGIAGVGAPDAAAAAAAAAKAAKFGAAGGFPGVGVGVGVPGVGVGVPGVG 298

Qy   308 ---VGVPGRAGVPGVPGVPGSAGIP----VVPGAGIPGAAVPGVVSPERAAAKAAAKAAKYGARPG 362
            ||||||||||||||||||||||||||:||  |||||||||||||| |||||||||||:||||
Db   299 VPGVPGVPGVGVGVPGVPGVPGVGVPGVSVGVPGAVSPAAAAKAAAKAAKFGARGG 358

Qy   363 VGVGGIPTYGVGAGGPGFG-------------------------VGVGGIPGVAG-VP----SVGGVPG 402
            |:||||||| |||| |                              |||:  |||| ||| ::||||
Db   359 VGIGGIPTPGVGPGGFPGIGDAAAAQAAAAAKAAKIGAGGVGALGGLVPGAGAIPGVPG 418

Qy   403 VGGGVPGVGISPERAQAAAAAAKAAKYGVGTPAABAAKAAAKAAKAQFGLVPGVGVAPGVGVGVGVAPG 462
            || | |||||||||||||||||||||||| |||||||||||||| ||||||||||||||||||
Db   419 -----PAAAAAKAAKAQFGLGPGVGVAPGVGVPG 459

Qy   463 VGVAPGVGLAPGVGVAPGVGVGVAPGVGVAPGVGISPGGVAAAKSAAKVAAKAQLRAAANGLGA 522
            ||||||||||||||||||||||||||||||||            ||||||||||||||||||||
Db   460 VGVVPGVGVAPGIGLGPG------------------GVIGAGVPAAAKSAAKAAAKAQFRAAAGLPA 508

Qy   523 GIPGLGVGVPGLGVGAGVPGLGVGVPGLGVGAGVPGPGA----------------- 542
            |:|||||||||||||||||||| || ::|||
Db   509 GYPGLGVGVGVPGLGVGVPGLGVGAGVPGPGA----------------------- 542

Qy   583 STPSSRVPGALAAKAAKYGAAVPGVLGGLGGVGIPGGVGAGPAAAAAAYAAAK 642
            |||||||||||||||||||||||:::|:|||||||||||||||||||||||
Db   543 ------VPGTLAAAKAKFGPGGVGALGGVGDLGGAGIPGGVAGVGP-AAAAAKFAAAK 594

Qy   643 AAQFGLVGAGLGGLGGVGGLG-VPGVGGLGGIPAAAAAAKYGAAGLGGVLGAGQFFPL 701
            ||||  | ||::||||||||:  ::||||||:|||||||||||||||||||:||||| ::
Db   595 AAQFGL--GGVGGLGVGGLGAVPGAVGLGGVSPAAAAKAAKFGAAGLGGVLGAGQFFPI 651

Qy   702 GGVAARPGFGLSPI----------------------TPPGGACLGKACGRKRK 731
            |||||||||||||                          ||||||::|||||||
Db   652 GGVAARPGFGLSPIFPGGAGGLGVGGKPFKPGGALGFPGGACLGKSCGRKRK 707
```

RESULT 7

```
ID   Q28097            PRELIMINARY;     PRT;   679 AA.
AC   Q28097;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN-CBEL2 (FRAGMENT).
```

```
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992576;
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J., Yoon K.;
RT   "Structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Cicila G., Yoon K., Rosenbloom J.;
RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30500.1; JOINED.
DR   EMBL; M14422; AAA30500.1; JOINED.
DR   EMBL; M19366; AAA30500.1; JOINED.
DR   EMBL; M19367; AAA30500.1; JOINED.
DR   EMBL; M19368; AAA30500.1; JOINED.
DR   EMBL; M19369; AAA30500.1; JOINED.
DR   EMBL; M19370; AAA30500.1; JOINED.
DR   EMBL; M19371; AAA30500.1; JOINED.
DR   EMBL; M22771; AAA30500.1; JOINED.
DR   EMBL; M22772; AAA30500.1; JOINED.
DR   EMBL; M22773; AAA30500.1; JOINED.
DR   EMBL; M22774; AAA30500.1; JOINED.
DR   EMBL; M22988; AAA30500.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   INTERPRO; IPR001459; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00959; MEVGALKINASE.
FT   NON_TER       1      1
SQ   SEQUENCE   679 AA;   57662 MW;   EB3C019E3BD7618D CRC64;
```

```
Query Match        63.8%;   Score 2413.5;  DB 6;   Length 679;
Best Local Similarity   57.7%;  Pred. No. 7.9e-109;
Matches 515;  Conservative  29;  Mismatches 78;  Indels 139;  Gaps  23;

Qy     28  GALGPGGKPLKPVPGGLAGAGLGAGLGAGLGAGPAVTPPGALVPGGAVADAAAYK-AAKAGA--  84
               | |||||| || |   |||||| ||||  ||||||| | |||||| |||||||||| ||||||
Db      1 -GLGGPGVKPZAKPGVGGGLVGPGLGLGAGLPG-AFPGAIVPGGPANGAAAYKAAKAAGAG  59

Qy     85 -GLGGVPGVGGLGVSAGAVVPQ---PGAGVKPGKVPGVGLPGSVTPCGVLP--GARPĞV  137
               |:||:| ||||||| |||||||| ||  ||||||||||||||||||||| |||||
Db     60 LGVGGSLGGVGGLSVSTGAVVPGLGAGVGAGVKPGVPGVGLPSVTPGGVLPGAGARPFGI  119

Qy    138 GVLPGVPTPGAGVKPKAPGVGGAFAGIPGVGPPGGPQPGVPLGGYTPIKAPKLPGGYGLPYTT  197
               ||||||||||||| ||||||||||||||||||||||  ||||||| ||||||||||||| |
Db    120 GVLPGVPTPGAGVKPKAPGGGGANFAGIPGVGPGPGGQQPGVPLGGYTPIKAPKLPAGYGLPYKT  179

Qy    198 GKLPYGYGPGGVAGAAGKAGYPTGTGVGPVGGPQARAAAKAAKPGAAAGVLP--GVGGAG  255
               ||||||||||||||||| |||| ||||||||||| || | | |||||||| |  |||||
Db    180 GKLPYGPGPGGVAGSAGKAGYPTGTGVGPQ-AAAAAKAAKALGAGGAGVLPGVGVGGAG  238

Qy    256 VPGVPGAGIPGIGGIAGVGTPAAAAAAAAAAKAAKYGAAAGL----VP--GGPGFG-PGV-  307
               :|| |||||||||||:|:||  |||||||||||||||||||       |    |||||| |
Db    239 IPGAPGAIPGIGGIAGVGAPDAAAAAAAAAKAAKFGAAGGFPGVGVPGVGVPGVGG  298

Qy    308 ---VGVPGAGVPGVGPGAGIP--VVPGAGIPGAAVPGVVSPEAAAKAAAKYGARPG  362
                   ||| |||| || |||| || || |:| ||| ||::|||||||||||||||
Db    299 VPGVGVPGVPGVGPGVGPGVGPGAGVPGVSPAAAAKAAAKAAKAKFGARGG  358

Qy    363 VGVGGIPTYGVGAGGPGFG---------VGVGGIPGVAG-VP----SVGGVPG  402
                ::  ||:| :|||||||:          || ::||||   :     :|||
Db    359 VGIGGIPTFGVGPGGFPGIGDAAAAQAAAAKIGAGGVGALGGGLVPGRPGAIPGVFG  418
```

```
Qy    403 VGGVPGVGVGGLSPEAQAAAAKAAKAKYGVGTPAAAAKAAKAAQFGLVPCVGVARPGVGVAPG  462
              |||||||||                      |||||| |||||| | ||:| |||:||:||
Db    419 VGGVPGVGI------------------------PAAAAKAAAAAQFGLGPSVGVARPGVGVVPG  459

Qy    463 YGVAPGVGLAPGVSVARPGVGVGVAPGVGPGIGPGGVAAAKSAAKYAAKAQLRAAAGLGA  522
              ||| ||||| |:||:| |                :  ||||||||| |||| |||||| |
Db    460 VGVVPGVGVAPGIGLGPG--------------GVIGAGVPAAAKSAAKAAKAQPRAAAGLPA  508

Qy    523 GIPGLGVGVPGLGVGAGVPGLGVGAGVPGFGAGADEGVRRNSLSPELREGDFSSSQHLP  582
              ||:|||||||||||||||||||||||||||||||||
Db    509 GVPGLGVGVGVPGLGVGVGVAGVPGFGAG------------------------------  543

Qy    583 SIPSSPRVPGALAAAKAAKAYGRAVPGVGGIGALGGVGIFGGVVGAGPAAAAAAKAAAK  642
              |:::   ::                 : |:| ||:  ||||  |||||| || |||||||
Db    544 -----PGGV-----------------GALGGVGDLGGAGIPGGVVAGVGP-AAAAAKAAAK  581

Qy    643 AAQFGLVGAAGLGLGVGGLG-VPGVGGLGGIPPAAAAKAAKYGAAGLGGVLGGAGQFPL  701
              |||| ||| || | | |||| ||| |||:||| |  |||||:|||||||||||||||| |
Db    582 AAQFGL---GGVGGLGVGGLGAVPGAVGLGGVSPARAAKAAKFGAAGLGGVLGAGQFFPI  638

Qy    702 GGVAARPGFGLSPI---------TPGGACLGKACGRKEK  731
                               |||||||||||||||
Db    639 GGGAGGLGVGGKPKPPGGALGALGFPGGACLGKSCGKKR  679
```

```
RESULT   8
Q28096
ID   Q28096          PRELIMINARY;       PRT;     666 AA.
AC   Q28096;
DT   01-NOV-1996 (TREMBLrel. 01, Created)
DT   01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE   ELASTIN-CBEL3 (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992576;
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J., Yoon K.;
RT   "structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C.C. Cicila G., Yoon K., Rosenbloom J.;
RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30501.1; JOINED.
DR   EMBL; M14122; AAA30501.1; JOINED.
DR   EMBL; M19366; AAA30501.1; JOINED.
DR   EMBL; M19367; AAA30501.1; JOINED.
DR   EMBL; M19368; AAA30501.1; JOINED.
DR   EMBL; M19369; AAA30501.1; JOINED.
DR   EMBL; M19370; AAA30501.1; JOINED.
DR   EMBL; M19371; AAA30501.1; JOINED.
DR   EMBL; M22771; AAA30501.1; JOINED.
DR   EMBL; M22772; AAA30501.1; JOINED.
DR   EMBL; M22773; AAA30501.1; JOINED.
DR   EMBL; M22774; AAA30501.1; JOINED.
DR   EMBL; M22775; AAA30501.1; JOINED.
DR   EMBL; M22988; AAA30501.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
FT   NON_TER       1      1
SQ   SEQUENCE   666 AA;   56435 MW;   BCB5B6263ZEE1B71 CRC64;
```

```
Query Match    59.8%;  Score 2262;  DB 6;  Length 666;
Best Local Similarity  64.9%;  Pred. No. 1.4e-101;
Matches 495;  Conservative 29;  Mismatches 83;  Indels 156;  Gaps 24;

Qy    28 GALGPGKPLKPVPBGGLAGLGAGLGRAFPAVTFPGALVPGGVADAAAYK-AAKAGA-- 84
         | :||| .|| ||||| || || | |||| | |||||
Db     1 GGLGPGVKPAKPGVGGLVPGGLVPGSLGAGLGALPG-AFPGALVPGGPAGAAAYKAAAKAGAAG 59

Qy    85 -GLGGVPGVGGLGVSAGAVVPQ----PGAGYKPGKVPGVGLPGVYPGGVLP--GARFPGV 137
         |:||| |||||||||||||||      ||||||||||||||||||||| |||||||
Db    60 LGVGIGGVGVGGLGVSTGAVVPQLGAGVKPGKVPGVGLPGVYPGVLPGVPGGVLPGAGARFPGI 119

Qy   138 GYLPGVPTGAGVYKPKAPGVGGAPAGIPGVGPPGGPQPGVPLGYPIKAPKLPGGYGLPYTT 197
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   120 GYLPGVPTGAGVYKPKAPGVGGAFAGIPGVGPGGQQPGVPLGYPIKAPKLP-------- 170

Qy   198 GKLPYGYGPGVGAGAGKAGYPTGTGVGPQARAAAAKAAAKFGAGAGVLP--GVGGAG 255
         ||||||||||||||||||||||||||||||||
Db   171 ------------------------------GVGPQ-AARAAAAKAAAKLPGVGVGGAG 204

Qy   256 VPGVPGGAIPGIGGIAGVGTVPARAAAAAKGVAAAGL----YP--GGPGFG-PGV- 307
         ||||||||||||||||||||||  |||    |||| : :||||
Db   205 IPGAPGAIPGIGGIAGVGVGGAPDAARAAAAAKARFGAAGGFPGVGVPGVGVPGVG 264

Qy   308 ---VGVPGAGVPGVGVPGAGIP---VVPGAGIPGAVPGVVSPEAAAKAAAKAKYGARPG 362
         |||||| | ||: ||| |||  |||||||||:|||
Db   265 VPGVGVPGVGVPGVGVGGVGVPGAVSPGAAAAKFGAGG 324

Qy   363 VGVGGIPTYGVGAGGPPGFG-------------------VGVGIPGVAG-VP---SVGGVPG 402
         ||:||||||| |||||                :|  ::   :|||
Db   325 VGIGGIPTPGVGPGGFPGIGDAAARAAQAAAAAKAAKIGAGGVGALGGLVPGGAPGAIPGVPG 384

Qy   403 VGGVPGVGISPEAQAAAAAKAKYGVGTPARAAAAKARAKAAQFG--LVPGVGVAPGVGVA 460
         |||||||
Db   385 VGGVPGVGI------------------PARAAAKAAKAAQFGIALLAFAGLGPGVGVA 425

Qy   461 PSGVAPGVGLAPGVGVAPGVGVAPGVGVAPGIGPGGVAAARASAAKVAAKAQLRAAAGL 520
         |||| || | ||||| ||  |:   |: :|||||| | |  ::|: |:|||
Db   426 PSVGVPGVVGPGLGVPGLGG----GVIGAGVPAAKSAAKSAAKAQPRAAAGL 480

Qy   521 GAGIPGLGVGVGAGVPGFGVGAGVTGGLGGALGGVGDGVVRRSLSPELREGDPSSSGCH 580
         |:| |||||||| |: |||||||:
Db   481 PAGVPGLGVGVGVGGLGVGVGVPGFGA----------------------- 516

Qy   581 LPSTPSSPRVPGALAAAKAAAAKGAAVPGVLGGLGALGGVGLIPGGVYGAGPGAAAAAKAA 640
         ||: |||:::||||||||    |     |||| | |||||||||
Db   517 ----VPGTLLAAAKARAKRFGPGGVRGALGGVGDLGGAGIPGVGVGCP-AAAAARRA 566

Qy   641 AKAAQFGLVGAAGLGGLGVGGLG--VPGVGGLGIPPARAAAKAAAAKFGAAGLGGVLGGAGQF 699
         |||||||||||||||| |||
Db   567 AKAAQFGL--GGVGGLGVGGLGVGAVPGAVGLGGVSPARAAAAAKAAKFGAAGLGGVLGAGQPF 623

Qy   700 PLGGVAARPGFGLSPI--------------FPGGACLLGKACGRKRK 731
         ||||||||||||||||
Db   624 PIGGGAGGLGVGGKPKPKFGGGALGLGFPGGACLLGKSCGRKRK 666
```

```
RESULT   9
Q28099
ID   Q28099;                PRT;   650 AA.
AC   Q28099;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN-CELI (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992575;
```

```
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J., Yoon K.;
RT   "Structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Cicila G., Yoon K., Rosenbloom J.;
RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30499.1; -.
DR   EMBL; M11422; AAA30499.1; JOINED.
DR   EMBL; M19366; AAA30499.1; JOINED.
DR   EMBL; M19368; AAA30499.1; JOINED.
DR   EMBL; M19369; AAA30499.1; JOINED.
DR   EMBL; M19370; AAA30499.1; JOINED.
DR   EMBL; M19371; AAA30499.1; JOINED.
DR   EMBL; M22771; AAA30499.1; JOINED.
DR   EMBL; M22772; AAA30499.1; JOINED.
DR   EMBL; M22773; AAA30499.1; JOINED.
DR   EMBL; M22774; AAA30499.1; JOINED.
DR   EMBL; M22775; AAA30499.1; JOINED.
DR   EMBL; M22988; AAA30499.1; JOINED.
DR   EMBL; M30010; AAA30499.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
FT   NON_TER        1      1
SQ   SEQUENCE   650 AA;   55373 MW;  CD21AEB3E9076AD7 CRC64;
```

```
Query Match    58.5%;  Score 2214;  DB 6;  Length 650;
Best Local Similarity  62.5%;  Pred. No. 2.8e-99;
Matches 484;  Conservative 26;  Mismatches 69;  Indels 196;  Gaps 22;

Qy    28 GALGPGGKPLKPVPBGGLAGLGAGLGRAFPAVTFPGALVPGGVADAAAAYK-AAKAGA-- 84
         | :||| .|| ||||| || || | |||| | |||||
Db     1 GGLGPGVKPAKPGVGGLVPGSLGAGLGALPG-AFPGALVPGGPAGAAAAYKAAAKAGAAG 59

Qy    85 -GLGGVPGVGGLGVSGAGAVVPQ----PGAGYKPGKVPGVGLPGVYPGGVLP--GARRPGV 137
         |:||| |||||||||||||||      ||||||||||||||||||||| |||||||
Db    60 LGVGIGGVGVGGLGVSIGAVVPQLGAGVKPGKVPGVGLPGVYPGVLPGVPGGVLPGAGARRFPGI 119

Qy   138 GYLPGVPTGAGVKRKAPGVGGAFAGIPGVGPFGGPQPGVPLGYPIKAPKLPGGYGLPYTT 197
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   120 GYLPGVPTGAGVKRKAPGVGGAFAGIPGVGPFGGPGQQPGVPLGYPIKAPKLP-------- 170

Qy   198 GKLPYGYGPGVGAGAGKAGYPTGTGVGPQAAAAAAKAAAKAAAKFGAGAGVLP----GVGGAG 255
         ||||||||||||||||||||||||||||||||
Db   171 ----------------------------GVGPQ--AAAAAAKAAAKAAAKAAKFGAGGVLPGVGVGGAG 204

Qy   256 VPVPGAIPCIGGIAGVGTVPGVSPGVGVAPGVGVAGL----YP--GGPGFG-PGV- 307
         ||||||||||||||||||||||  |||    |||| : :||||
Db   205 IPAPGAIPCIGGIAGVGPDAAAAAAAKAAAKAAKFGAGGFPGVGVPGVGVPGVG 264

Qy   308 ---VGVPGAGVPGVGVPGVGVPGAGIP---VVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGARPG 362
         |||||| | ||: ||| |||  |||||||||:|||
Db   265 VPGVGVPGVGVPGVGVPGVGLGVGVPGAVSPAAAKAAAKAAKAAAKAAKFGARGG 324

Qy   363 VGVGGIPTYGVGAGGPPGPG-----------------VGVGIPGVAG-VP---SVGGVPG 402
         ||:||||||| |||||                :|  ::   :|||
Db   325 VGVGGIPTPGVGGGFPGPGDAAAAQAAAAAKAAKIGAGGVGALGGLIVPGPGAGAIPGVPG 384

Qy   403 VGGVPGVGISPEAQAAAAAKAKYGVTPAAAAAKAAKYGVTPAAAKAAAKAQPGLVPGVGVAPGVVAPG 462
         |||||||
Db   385 VGGVPGVGI-------------------PARAAAKAAAKAQPGLPVPGVGVAPGVGVVAPG 425

Qy   463 YGVARPGVGLAPGVGVAPGVGVAPGVGVAPGGLRAAAGLGA 522
         ||:||||||
Db   426 YGVVPGVGVAPGIGLGPG---------GVIGAGVPAAAKSAAKARAAAQFRAAAGLPA 474

Qy   523 GIPGLGVGVGVPGLGVGANGVPGLGVGAGVPGFGAGADEGVRRSLSPELREGDPSSSQHLP 582
```

```
Db   475  GVPGLGVGVGVPGLGVGVPGLGVGAGVPGFGA------------------------ 508
          :!|||||||||||  ||||||||||||||||
Qy   583  STPSSSPRVPGALAAAKAAKYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAAKAAAK 642
          ||||||  ||||||||||:||:  | ||||||  || ||||||||||: -AAAAAKAAAK 560
Db   509  ------VPGTLAAAKAKPGPGGVGALGGVGDLGGAGIPGGVAGVGP-AAAAAKAAAK 560
Qy   643  AAQFGLVGAGLGGLGVGGLGVPGVVGGLGGIPPAAAAAKAAKYGAAGLGGVLGGAGQFPLG 702
          |: !||||||||||||          |: ||||||||||||||       !|:|
Db   561  AAQF------------------------GVSPAAAAAKAAKFGAAGLGGVLGAGQPPFIG 595
Qy   703  GYAARPGPGLSPI--------           FPSGACLGKACGRKRK 731
          |||||||||||||||
Db   596  GVAARPGFGLSPIFPGGAGGLGVGGKPKPFPGGALGALGFPGGACLGKSCGRKRK 650
```

```
Query Match         33.2%;  Score 1258.5;  DB 4;  Length 258;
Best Local Similarity  88.5%;  Pred. No. 7.3e-54;
Matches  253;  Conservative  0;  Mismatches  0;  Indels  33;  Gaps  1;

Qy   446  GLVPGVGVARGVGVARGVGVARGVGVARPGVSGVARPGVGVARPGVGVAPGIG2GGVAAAKSA 505
          || ||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   6    GLVPGVGVARGVGVARGVGVARPGVGVARPGVGVARPGVGVAPGIGPGGVAAAAKSA 65

Qy   506  AKVPAKAQLRAAAGLAAGIPGLGVGVPGLGVGVPGLGVGVPGPGVGADEGVRRS 565
          ||||||||||||||||||||||||||||||||||||||||
Db   66   AKVPAKAQLRAAAGLGAGIPGLGVGVPGLGVGVPGLGVGVPGFGA---------- 116

Qy   566  LSPELREGDPSSSQHLPSTPSSPRVPGALAAAKAAKYGAAVPGVLGGLGALGGVGIPGGV 625
          |||||||||||||||||||||||||||||||||||||
Db   117  -------------VPGALAAAKAKPGPGGVGALGGVGDLGGAGIPGGV 152

Qy   626  VGAGPAAAAAKAAAKAAAKAAAQFGLVGAGLGGLGVGGLGVPGVVGGLGGIPPAAAAAKAAKYG 685
          ||||||||||||||||||||||||||||||||
Db   153  VGAGPAAAAAKAAAKAAAQFGLVGAGLGGLGVGGLGVPGVVGGLGGIPPAAAAAKAAKYG 212

Qy   686  AAGLGGVLGGAGQFPLGGQVVARPGPGLSPIFPGGACLGKACGRKRK 731
          |||||||||||||||||||||||||||||||
Db   213  AAGLGGVLGGAGQFPLGGQVVARPGPGLSPIFPGGACLGKACGRKRK 258
```

```
Query Match         20.1%;  Score 759.5;  DB 5;  Length 988;
Best Local Similarity  35.4%;  Pred. No. 2.2e-29;
Matches  261;  Conservative  7;  Mismatches  321;  Indels  149;  Gaps  24;

Qy   5    GAIPGGVPGGVFYFGAGLGALGGALGPGGKPLKPVPGGLGAGLGAGLGAFPAVTPFGA 64
          || |||| |   ||| || ||||  | :|  |  | !|||| ||  |  || |
Db   266  GAGAGGAGGYGRGAGAGAGAAGAAGAGAGAGAGAGY----GGQGGYGAGAGAGA-AAAAGGA 320

Qy   65   LVPGGYADAAAAYKAAKAGAGAGLGGPGVGVGGLGVSAGAVVPQPGAGYKPGKVPGVGLPGVY 124
          || ||||  |   !||| ||| |||| |||  | | ||   |
Db   321  GGAGGYGRGAGAGAGAGGYGGQGGYGAGAGAAGAAGAGAGSGGAGYG--------- 375

Qy   125  PGGVLPGARFPGVGVLPGVPTGAGVKPKAPGVGGAPAGIPGVGPFPGGPQPGVPLGYPKKA 184
          ||  |||       |  |  |  |:||
Db   376  ----RGAGAGAGAGAGA----------GAGSYGG-------- 397

Qy   185  PKLPGGYGLPYTTGKLPYTGPPGGVYAGA---AGKAGYPTGTVGPGVQAAAAAAKAAAKFG 241
          ||||
Db   398  ---QGGYGA---GAGAGAAAAGAGAGYGRGAGAGAGAAGAAGAGAAAAAGAGAG 445

Qy   242  AGAGVGGQGGYGAGAGAGAAGAAAAAAGAGAGGAGYGRGAGAGAGAAGLVPGG 300
          : |  |    |  ||  |    |   |  |   | ||| |:|  |||  |
Db   446  GAGVGGQGGYGAGAGAGAAGAAAAAAGAGAGGAGYGRGAGAGAGAAAKYGAR-QSG 503

Qy   301  PGPGPGVYGVPGAGYGVGYGVARPGVGVARPGVVPGGAGIPGAAVPGVVSPEAAAKAAKYGAR 360
          ||||||||||||||||||||||||||||||||||||||||||||
Db   504  YGAGAGAAAAGAGAGGAGGYGRG-----AGAGAGAAAGAGAG 554

Qy   361  ----PGVVGGIPPTYGVGAGGFPGFRGVSVGGIPGVVGGVPGGVPGVVGGVPCVGISPRAQ 416
          :   | || | | || ||  | | |  | | | |  || ||  || || | |
Db   555  GGYGAGAGAGAAAAGAGAGYGRGAGAGAGYGRGAGAGGYGGQGGY-GAGAGAGAA 613

Qy   417  AAA---AAKAAKYGVGIPARAAAKAAAKAAQFGLVPGVGVAPGVGVAPGVGLAP 473
          |||  : ||| | |   ||||     |  |    |||||  | |  | |  |:|
Db   614  AAATGAGGAGGYGRGAGAGAGAAGAGAGTGGG--AGYGGQGCYGAGAGAGAAAGAGA 671

Qy   474  GVGVAPGVGVABGVGVABGVGVARPGIPGGVAAAAKSAAKVRAKAQLRAAAGLGAGIPGLGVGVGV 533
          || || ||| |||  | |  ||  || || | |   |  || ||| ||||||||||| |||||||
Db   672  G-GAGYGRGAGAGAGAAAGAGAG-----------AAGAGAGGYG-GQGG 710

Qy   534  PGLGVGAGVPGLGVGAGVPGFPGAGADEGVRRSLSPELREGDPSSSQHLPSTPSSPRVPGA 593
          || |||||||||||||||  ||| |   |     |     | |     | | |
Db   711  YGAGAGAG-----AAAAGAGAGGAAGYSR-------------------GG 737

Qy   594  LAAAKAAKYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAKAAA---KAAQFGLV 649
          |:||  | | |  || |||   |  | :|||   ||||  ||   |:|||
Db   738  RAGAAGAGAAA-GAGAGAGAGGYGGGIPPAAAAKAAKYIGAAGLGGVLGGAGQFPLGGVAARPG 794

Qy   650  GAAGLGGLGVGGLGVPGVGGLGGIPPAAAAKAAKYIGAAGLGGVLGGAGQFPLGGVVAARPG 709
          || |||||| | |||  | |  || | ||| || !|||
Db
```

```
Db   795 AAAGAGAAGAGGAGAGGYGGCGGYGAGGAGAGAAAGAAGAGRGGYGRGAG--AGGYGGGG 851

Qy   710 PGLSPIFPGGACLGKACG 727
          :: :|
Db   852 YGAGACAGAAAAAGAGAG 869

RESULT 12
ID   Q9U617
AC   Q9U617;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PERICARDINE.
GN   PRC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON R;
RA   Chartier A., Astier M., Zaffran S., Semeriva M., Gratecos D.;
RT   "Characterization of an extracellular matrix component specifically
RT   expressed in the basal lamina of the dorsal vessel in Drosophila
RT   melanogaster.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF203342; AAF13280.1; -.
SQ   SEQUENCE   1729 AA;  164661 MW;  DA9B1B7FAB5CDEDB CRC64;

Query Match          20.0%;  Score 756;  DB 5;  Length 1729;
Best Local Similarity 32.9%;  Pred. No. 5.4e-29;
Matches 286;  Conservative  50;  Mismatches 302;  Indels 232;  Gaps 55;

Qy    1 GGVPGAIPGGVPGGVVFYPGAGLAGLGGGALGPGGGKPLKPVPGGLAGAGLGAFPAV- 59
Db  132 GGVGGGQTGAGQPGYGSQPGIG---GQTATGQPGYGSQLGVGVGAQAGAG-QPGYGAQPGVG 186
Qy   60 TFPGALVPGGVADAAAYKAAKAGAGLGGVPGVGGLGVSAGANVVPQPGAGVKP--GKVPG 117
Db  187 AQTGAGQPGYGSQTGIGGGQTGAGQPGYGSQPGIGG--QTGA-QPGYGSQPGIGGQTG 241
Qy  118 VGLPGV--YPG-GVLPGARFPGVGVLPGV--PTGAGVKP--KAPGVGG-AFAGIPGVGP 168
Db  242 AGQPGYGSQPGIGGQTGAGQPGYGSQPGIGGQTGAG-QPGYGSQPGIGGQTGAGQPGY-- 298
Qy  169 FGGQPGV----PLGYPIKAPKLPGGY----GLPYTTGKLPYGYGP----GGVAGAGKA 216
Db  299 --GTQGIGGQTGAGQP------GYGSQPGIGGQTGAGQPGYGSQPGIGGQTG-AGQP 347
Qy  217 GYPTGIGVPQAAAAKAAAKFPGAGAAGVLFGVGGAGVPGVP--GAIPGIGGIAGVGT 274
Db  348 GYGSQPGIGGQTGA------------GQPGYGTQPGIGGQTGAGQPGYGTQPGIGGQTGPGCQ 397
Qy  275 PAAAAAAAAAKAAAKYGAAAGL----VPGGCPGPG--PGYVVPGAGYPGVGV-PGAGIPW 327
Db  398 PG--------YGSQPGIGGQTGAGQPGYGTQPGIGGQTGAGQPGYGTQPGIGGQAQT 444
Qy  328 VPGAGIPG-RAVPGVVSPERAAAAAAAKAAAKYGARPGVGV--GGIPTV----GVGA--G 376
Db  445 --GAGQPGYGTQPGI----GGQTGAGQPGYGTQPGIGVQTGAGQPGYGSQPGIGAGQTGA 497
Qy  377 GFPGKG--YGVGGIPGVAGVVSVGGVPGVG-----GVEGVGISPEAQAAAAAKAAKYGVG 429
Db  498 GQPGYGSQPGIGGQTGA-GYGSQPGIGGQTGP-GIGAQTGAGQPGYG- 554
Qy  430 TPAAAAAKAAKAAQFGLVPGVG-----VAPGVVAPGVGVAPGVGVLAPGVGVAPGVGV- 483
Db  555 SQPGIGGQTGAGQPGYGSQPGIGGQTGAGQPGYGTQPGIGGQTGAG-QPGYGSQPGVGTQ 613

Qy  484 ---APGVGYVAPGIG------PG----GVAAAAKSAKVAAKAQLRAAAGAGAGIPGL 527
Db  514 IGAGQPGYGSQPGIGGQTGAGQPGYGSQPGIGGQT-GAAQPGYGTQPGYGAQTGTGQPGY 672
Qy  528 G------VGVGVPGLG-------VGAGVPGLG--------YGAGVPGFGAGADEGVR- 563
Db  673 GAQPGIGGQTGAGQPGYGIGRGPGIGGQTGPGQPGYGTGPGVGVGTQTGTGQPGYGAQPGIG-- 730
Qy  564 RSLSPELREGDPSSSQHLPSTPSPSPRVPGALAAAKAAKYGAAVPGVLG--------GL 613
Db  731 ------GQSGAGQ--PGYGSQPGIGGQTGGGQPG-YGSQIGGCQTGAGQPSYGSQPGV 778
Qy  614 GALGGVGIPGGVVGAGPAAAAAKAAAKAAQPGLVGAAGLGGL-------GVGLGVPG 666
Db  779 GAQNGGGQPG--YGTRPVTGGQTG-----AGQPGYGGQTGVGGSPGFLTQGIGGISGP- 830
Qy  667 VGGLGGIPPAAAAKAAKYGAAGLGGV--------LG--GAGQPPLGG----------- 703
Db  831 IGGKVGGGQSEAAKPGYWAQPGIGGPSRYGSQPGIGDQTGAGQSGYGGQPGISAGQTGGQ 890
Qy  704 -------VAARGFGLSP-------IFPGG 719
Db  891 PGYGGQATISGLRGYTQPGIGALTAVPGG 920

RESULT 13
ID   Q9NHW2
AC   Q9NHW2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   FLAGELLIFORM SILK PROTEIN (FRAGMENT).
GN   FLAG.
OS   Nephila madagascariensis.
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC   Araneomorphae; Entelegynae; Araneoidea; Tetragnathidae; Nephila.
OX   NCBI_TaxID=115969;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hayashi C.Y., Lewis R.V.;
RT   "Molecular architecture and evolution of a modular spider silk protein
RT   gene.";
RL   Science 287:1477-1479(2000).
DR   EMBL; AF218623; AAF36091.1; -.
FT   NON_TER   1884   1884
SQ   SEQUENCE   1884 AA;  148738 MW;  27B6F45339FD20A5 CRC64;

Query Match          19.7%;  Score 744.5;  DB 5;  Length 1884;
Best Local Similarity 33.6%;  Pred. No. 2.1e-28;
Matches 286;  Conservative 27;  Mismatches 354;  Indels 185;  Gaps 43;

Qy    2 GVPGAIPGGV-PGGVVFYPGAGLGALGGGALGPGGKP-KPVPGGLAGAGLGAGLGAFPAVT 60
Db  751 GPGGAGPGGSGPGGYGPGSSCPGGYGPGGIGPGGYG----PGGAGPGGAGPG-GAGGSGA 805
Qy   61 FPGALVPGGVADAAAAY-----KAAKAGAGLGGVPGVGGLGVSAGA------- 101
Db  806 GPGGAGFGSGAGRGGSGRNSGVSRGGAGRGGAGGAGGSGAGGSGGGTTIIB 865
Qy  102 -------VFPQPGA-GVKPG--KVPGVGLPGVTPGGVLPGGARFPG----136
Db  866 DLDITVDGANGPIISEELTIGGAGGGVPGGSGPGGVPGGSGPGGGVGVGPGGGVG 925
Qy  137 ------VGVLPG--VPTGAGVKPKAP-GVGGAFAGIPGVGP--FGGPQ------- 173
Db  926 SGGSGPGGVGPGGSGPGGVGSGPGPGGVGGPYGPGGSGPGGPGGAGG 985
Qy  174 ------EGVPLGYP--------1KAPKLPGGYGLPYTTGKLPYG------- 203
Db  986 AGGSYGPGGPYG-PGGSSCGPGGAGGPYGPGGAGPYGPGPGPGGSGPGGAG 1043
```

```
Qy      204  --YGPGGVAGA---------AGKAGY-PTGTGVGPQAAAAAAARKAAKFGAGAGAGYLPGV   251
                 |||| ||    :     |  ||||    || |::  |   || ::||  ::
Db     1044  GPYGPGPGPGGAGPGGYGPGGEGPGGYGPGGEGPGGYGPGGAGPGGYGPGGAGPGGYGPGG  1103

Qy      252  GGAGVPGVPGAIPGIGIAGVGTPAAAAAAAAKKAAKYG-AAAGLVPG--GPGEGPGVV     308
             |  | ||  : |  | :                |   |  |::  |    ||  |||
Db     1104  PGPGGSGYGPGGAGPGGYGPGGYTGPGGSAPGGAGPGGYGPGGSGPGGYGPGGSGPGGA  1163

Qy      309  GVPEGAGVPGVGVPGAGIPVVPGAGIPGGAAVPGVVSPEAAAKARKAAKYGARP-GVSVGG   367
             |||  |  | | :| || | :  ::|  |  |:   :||| |  ||| |     |||:||
Db     1164  GPGGAGPGGAGPGGAG---PGGAGPGGAGPGGAGPGG--GAGPGGVSTGG   1217

Qy      368  :IPTYGVGAGGRPFGPGVGVGGI------PGVAGYPSVGGVPGVGGVPGVGISPEAQAAAA   420
                | | :|| :||  | : |:|     | ||  ||: |  | |:  |||:|  |  |
Db     1218  LGRGGAGRGGAGRGGAGRGGAGRGGYGPGGAGGSGGTTVIED   1277

Qy      421  AKAAKYGVGTPAAAAAAAAAAKKAAQFGLVPGVCVGVARPGVGVTAPGVVRAPGVGLAPGVGVAPG   480
             || ||| ||||                 || :|:| ||  |  |  | :  | | |  ||
Db     1278  LDITIDGADGPTIISEELTISGAG------GSGPG-GAGPG-GVGPG-GSGPG-GVGPG  1328

Qy      481  VGVAPGVGVAPGIGPGGVAAAAKSAAKVAAKAQLRAAAGLAGAGIPGLGVGVGPGLGVGA   540
             |  || | || |  ||                      | |:|   || |  |  ||
Db     1329  -GSGPG------GVGPGGAGGPGGAGXGP   1366

Qy      541  GVPGLGVGAGVPGFGAGADEGVRRSLSPELRBGD---PSSSOHLFSTPSSPRVPGALA--   595
             | ||||  | |  : |                :        ::|:  | | | |:
Db     1367  GGSGGPGGAGGP--YGPGG-EGPGGSGGPYGPGGEBPGGAGGPYGPGGAGGGP   1424

Qy      596  ---RAKAAKYGAAVPGVGVLGGLGALGGVG IPGGVVGAGPAAAAAAAAKARAAQPGLVGAA   652
                  |||| |  |||  :  |  |   :  | |||  |:|  ||||||||| |   || ||
Db     1425  YGPGGAGGPYG---PGGAGGPYGPGGAGGPYGPGGAGGPYGPGGAGGPYG---GVAARP   1480

Qy      653  GLGGLGVGLGVPGVGVGGLGGIPPAAAAKAAKYGAAGLGG-GVLGGAGQFPLG---GVAARP   708
             | :| |:|  |||   | |  |    || || :|  | |  |  |     |     |||||
Db     1481  GPGGTGPGGGYG-PGGAGPGGYGPGGSGPGG-YGPGGSGPG-GGSYGPGPGGSGPG   1538

Qy      709  GFGLSPIFPGGA   720
             | | |:|||||
Db     1539  GYGPGGGSGPGGA   1550
```

```
RESULT 14
Q9NHW4
ID   Q9NHW4            PRELIMINARY;      PRT;   2249 AA.
AC   Q9NHW4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   FLAGELLIFORM SILK PROTEIN (FRAGMENT).
GN   FLAG.
OS   Nephila clavipes (Orb spider).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC   Araneomorphae; Entelegynae; Araneoidea; Tetragnathidae; Nephila.
OX   NCBI_TaxID=6915;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hayashi C.Y., Lewis R.V.;
RT   "Molecular architecture and evolution of a modular spider silk protein
RT   gene.";
RL   Science 287:1477-1479(2000).
DR   EMBL; AF218621; AAF36090.1; -.
FT   NON_TER      1        1
FT   NON_TER   2249     2249
SQ   SEQUENCE   2249 AA;  174871 MW;  88CB8168A147CDA0 CRC64;
```

```
Query Match         19.1%;   Score 723;   DB 5;   Length 2249;
Best Local Similarity  35.6%;   Pred. No. 2.6e-27;
Matches  268;  Conservative  22;  Mismatches  325;  Indels  138;  Gaps  43;

Qy        2  GVPGAIPGGV-PGGVPYPGAGLGALGGALGGALGGALGGCKPLKPVPGGLAGAGLGAGLGAFPAVT   60
                : :| |||  ::: ||| | |   |    ||||     | |  |
Db     1541  GPGGSGPGGYGPGGSGPGGAGPGGVGPGPGGAGPG---PGGAGPGGAGPG-GAGPGGA  1595
```

```
Qy       61  FPGALVPGVADAAAAYKAKAGAGLGGVPGVGGLGVSAGAVPQ----------   105
                |||  | :  | | |  ||  | |:| ||: : | | | ||:
Db     1596  GPGGAGPGGAGPGGAGPGGAGPGGVLPGVTPGGVLRGARPPGVGLPGVTTGAGVKPKAPGVGGA  1655

Qy      106  ---------PGAGVKPKGKVPGVGLPGVGVLPGVTPGGVLRGARPPGVGLPGVGVTGAGVKPKAPGVGGA   159
                       || || ||     :                      |  || | |  | |:|
Db     1656  SEELTISGAG---GSGPG'---GAGPGGVGPGGSGPG-GVGPGGSGPGGSYGPGGSGPGGV  1707

Qy      160  FAGIPGVGFF--GGPGPGVFLGTPIKAPKLPGGYGLPTTGKLPYGYGPGGYAGAAGKAG   217
             : | ||:       |  ::   |  || | | : |   || ||||||::|  | |:|
Db     1708  GPGGAG-GPYGPGGSGPGGAGG----------AGGPGGAYGPGG---SYGPGGSGPGGAGG   1755

Qy      218  --TPTGTGVGPQAAAAAAAAAKKAAKFGAGAGAGVLPGVGAGVPGVPGAIPGIGGIAGVGTP   275
               ||  | | |     | ||  ::||||     |    || ||  |   |  || | ||
Db     1756  PYGPGGEGPGGAGGPYGPGGAGPGGPGGSGPGGGSYGPGGSYGPGGAGGPGP   1815

Qy      276  AAAAAAAAKAKYGAAAGLVPGGPG---EGPGVVPGVGPGAGYPGVGYGPGGEPGGSYGPGGAG---PGGYG   333
             ||||||||||  |   |||            | | | || | | ||| |         || |
Db     1816  ------------GGPYG--PGGEGPGGAGGPYGPGGYGPGGSGPGGAG   1859

Qy      334  FGAAVPGVVSPEAAAKAAAKAAKYGARPGVVGGIPTYGVGARGGPFPGPGGVGVGGIPGVTAG   393
                | ||:|   |  | | | ||  |     |:                |  || |  :| |||
Db     1860  PGSSGPGGYGPGGSGPGG-------YGP-GGSGPGG---YGPGGSGGPGGYGPGGSGPGGSG   1909

Qy      394  VPSVGGVPGVGGYVPGVGISPEAQAAAAKAAAKYGVGTPAAAAAAAKAAAAQFGLVPGVGV   453
             | | || ||||    | | | |  | || || |     |||||||||||||| | | |||
Db     1910  PGGYG--PG-GSGPG-GVGPGGSGPGSGYGPG---------------GSGPG'GF   1948

Qy      454  ARPGVTAPGVGVAPGVGLAPGVVARPGVVARPGVGVTAP-GIGPGGVAAAAKSAAKVRAKA   512
             | |: :| | |  | |:| | | | || ||  |                 ||| |
Db     1949  GPG-GFGPG-GSGPG-GYGPG-GSGPG-GAGPG-GVGPGGGPGGAGPGGAGPGG   2002

Qy      513  QLRAAAGLGAGIPGLGVGVGPGLGVGAGVPGFCAGADEGVR----RSLSP   568
                   | | | | | |:|  |  | | | | |  |   |:    |:|:|
Db     2003  AGPGGAGPGGAGPG-GAGPGGAGPGGAGGAG-GAGGSG-GAGGSGGPTIIEDLDITI   2059

Qy      569  ELRBGDPSSSOHLPSTPSSPRVPGALAAAAKAAKYGAAAVPGVGGLGALGGVGIPGGVVGA   628
             |:  | |:: :|  |  |      ||  |   |  ||   | ||| | |  | |||| |
Db     2060  DGADGPTIISEELTISGAGGSGPG-----------------GAGP3GGVGPGGSGPGGAGGAG-PG--'GS  2105

Qy      629  GPAAAAAAKAAKRAAQPGLVGAAGLGG-LGVGGLGVPGVGGLGGITPAAAAKAAKYGAA   667
             ||||||||| | :||  ||  |       |  | | || | | |  | || || | |||
Db     2106  GPGGVGPGGSGA-----GGVGPGGAGGPYGPGGSGPGGAGGAGG-PGGAYGPGGSTYGPG  2158

Qy      688  GLGGVLGGAGQPFLGGVAARPGFGLSPIFPGGA   720
             | ||:| |  | ||   |:||  | | |||||
Db     2159  GSGGPGGAGGPYGPGGEG---PGGAGGPYGPGGA  2189
```

```
RESULT 15
044359
ID   044359            PRELIMINARY;      PRT;   907 AA.
AC   044359;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   FLAGELLIFORM SILK PROTEIN (FRAGMENT).
GN   FLAG.
OS   Nephila clavipes (Orb spider).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC   Araneomorphae; Entelegynae; Araneoidea; Tetragnathidae; Nephila.
OX   NCBI_TaxID=6915;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98153262; PubMed=9480768;
RA   Hayashi C.Y., Lewis R.V.;
RT   "Evidence from flagelliform silk cDNA for the structural basis of
RT   elasticity and modular nature of spider silks.";
RL   J. Mol. Biol. 275:773-784(1998).
DR   EMBL; AF027973; AAC38847.1; -.
DR   INTERPRO; IPR000087; -.
FT   NON_TER      1        1
```

SQ   SEQUENCE   907 AA; 72392 MW; A0864EC708740A00 CRC64;

Query Match   18.4%; Score 698; DB 5; Length 907;
Best Local Similarity 33.2%; Pred. No. 1.8e-26;
Matches 261; Conservative 29; Mismatches 309; Indels 186; Gaps 42;

```
Qy      2 GVPGAIPGGV-PGGVFYPGAGLGALGGGALGPGGK------PLKPVPGGLAGAGLGAGLG 54
            |||  || ||| || :|| ||  ||||| ||| | :      ||| ||| ||| | ||||
Db    131 GPGGSGPGGYGPGSGPGGAGPGGVGPGGFGPGPGGAGPGGAGPGGAGPGGAGPG-G 189

Qy     55 APPAVTPPGALVPGGVADAAAYKAAKAGAGLGGVPGVGGLGVSAGAVPQ--------- 105
            | ||| || |  ||| | | ||||||| ||  : |    |||||| | :
Db    190 AGPGGAGPGGAGPGGAGPGGAGGAGGSGGAGGSGGTTIIEDLDITIDGA 249

Qy    106 --------PGAGYKPGSKYPGVGLPGVYPGGVLPGARFPGVGVLPGVPTGAGVKPKA 153
                    | | :|| :|  ||| || ||   |||  | :| || : :|
Db    250 DGPITISEELTISGAG---GSGPG----GAGPGGVGPGGSGPG-GVGPG---GSG- 293

Qy    154 PGVGGAFAGIPGVGFFGGFGGPQPGVPLGYPIKAPKLPGGVGLPYTTGSKLPYGYGPGVAGRA 213
            |||     |||  |  : : | | | | |||| :| |  ||||  | || |||  :|
Db    294 PG------GVGP-GSSGPG-GVG--------PGGAGGPYGPG---GSGPGGAGGAG 330

Qy    214 GKAG-YPTGTGVGPQAAAAAAAKAAKYGAGAGVLPGGVGGAGVPGVPGAIPGIGGIAGV 272
            || |  ||||| :||||| | | |||| | :||   ||  | :| ||| :| ||| ||
Db    331 GPGGAYGPGGSYGP-GGSGGPGGAGGPYGPGG---EGPGGAGGPYGPGGAGGPYGPGGA 385

Qy    273 GTPAAAAAAAAAKAAKYGAAAGLVPGGEPG-FGPGVVGVP-GAGVP----GVGVPGAGI 325
            ||        | :|  :  : :    :|     ||| | :| | :|    ||| || :||
Db    386 GGP-----------YGPGGEGPYGPGGAGGPYGPGPGGEGPGGAGG 431

Qy    326 PVVPGAGIPGAAVPGVVSPEAAAKAAKAAKYGARPGVGVGGIPTYGVGAGGFPGFGVGV 385
            ||| ||| | |   |||  |||  | : :|      ||| | : |||||| || |||||
Db    432 PYGPGGVGPGGSGPGGYGPGSSGPGG------YGP-GGAGPGGYGPGGSGPGGYGPGGSGP 485

Qy    386 GGI-PGVAGVPSVG-GVPGVGGVPGVGISPEAQAAAAKAAAKYGVGPPAAAAKAAAKAA 443
            ||   | |   : |  |  |||| | |   :||  | | | |||  |   ||  | |||
Db    486 GGYGPGSGPGGYGPGSPGGYGSGGAGPGGYGPGCGSGPGGYGPG----- 531

Qy    444 DFGLVPGVGVAPGVGVAPGVGVADGVGLAPGVGVAPGVGVAPGIGPGGVAAAK 503
            : |  | |  |  ||   |  | | ||| | ||| |  |||  || ||| | :|
Db    532 --GSGPG-GYGPG-GYGPG-GSGPG-GYGPG-GSGPGGSGPGGSGPGGYGPGSGS 583

Qy    504 SAAKYAAKAQLRAAAGLGAGIPGLGVGVGVPGLGVGAGVPG--------- 553
            | | | :|| ||| :|  ||  |  ||  | |  |||  | ||
Db    584 GPGGYGPGSGPGGYGPGGSGPG--GYGPGGSGAGGTGPGAGGAGGAGGSGGGAGGSGG 640

Qy    554 ----FGAGADEGVRASLSPELRE------------GDPSSSQHLPSTPSSPRVPGALAAAKA 599
               |  |      ||: |  |                |  ::| | :| |: |  | |  :| |
Db    641 AGGSGGAGGSGGVGGSGGGTTIIEDLDITIDGADGPITISEELTISGAGGSGPGGAGPGGV 700

Qy    600 AKYGRAVPGVLGGLALGGVGIPGGVVGAGVPAAAAKAAAAKKAAQFGLVGAAGLGGLGV 659
            |: |  | | || |  ||||||: |   |     | | : :|  |   |  :|| ||
Db    701 GPGGSGPGGYGPGSGPGGVG--PG---GSGPGGVGG-----------GGSGPGGVGP 742

Qy    660 GGLGLVPGVGLGGIPPAAAAKAAKYGAGLGGVLGGAGQ---FPLGGVAARPGFGLSPI 715
            ||| | :| :  |   |||| | | ||  ||  || |:|     |  || | || :|
Db    743 GGYG-PGGSSGGFYGPGGSSGGEYGPGGSGGFYGPGSTYGSGGGYG---- 791

Qy    716 FPGGA 720
            ||||
Db    792 -PGGA 795
```

Search completed: April 24, 2001, 16:40:24
Job time: 414 sec

```
                GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    April 24, 2001, 16:38:14 ; Search time 115.25 seconds
                                  (without alignments)
                                  99.695 Million cell updates/sec

Title:          US-09-340-736-2
Perfect score:  1017
Sequence:       1 FPGFGVGVGGIIPGVAGVPGV.........GVAPGVGVAPGVGVA.PAIGP 201

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0:5

ched:      390729 seqs, 57263235 residues

Total number of hits satisfying chosen parameters:       390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_0401:*
        1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
        2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
        3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
        4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
        5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
        6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
        7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
        8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
        9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
       10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
       11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
       12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
       13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
       14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
       15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
       16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
       17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
       18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
       19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
       20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
       21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
       22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

                        SUMMARIES

Pred: No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1017	100.0	201	19	W46316	Non-natural polype
2	709.5	69.8	698	20	Y01302	Human tropoelastin
3	709.5	69.8	698	21	Y59069	Amino acid sequenc
4	709	69.7	660	20	Y01303	Human tropoelastin
5	707.5	69.6	712	21	B08630	Amino acid sequenc
6	707.5	69.6	730	21	B08631	Fusion protein com
7	700	68.8	730	19	W46315	Human elastin cont
8	700	68.8	731	21	Y59060	Amino acid sequenc
9	700	68.8	733	15	R56653	Synthetic human tr
10	700	68.8	733	20	Y01301	Amino acid sequenc
11	683.5	67.2	571	21	Y69071	Amino acid sequenc

```
     12   670   65.9   515   21   Y69135   Amino acid sequenc
     13   531.5 52.3   745   20   Y31682   Alanine containing
     14   455.5 44.8   148   20   Y31681   Alanine containing
     15   446.5 43.9   294    9   P82484   Tropoelastin. Gal
     16   401   39.4   988   16   R80253   Polymer SELP7.  Sy
     17   397.5 39.1   972   16   R80255   Polymer SELP4.  Sy
     18   392   38.5  1056   16   R80254   Polymer SELP3.  Sy
     19   391.5 38.5  1024   16   R80256   Polymer SELP5.  Sy
     20   390   38.3   832   16   R80252   Polymer SELP8.  Sy
     21   382   37.6   936   16   R80251   Polymer SELP0.  Sy
     22   379.5 37.3  1412   19   W53519   Amino acid sequenc
     23   379.5 37.3  1413    9   P82957   ESBI protein compr
     24   379.5 37.3  1413   18   R41008   EBSI multimeric pr
     25   379.5 37.3  1413   18   W26343   EBSI synthetic ela
     26   379.5 37.3  1464   21   Y78278   EBSI protein seque
     27   376.5 37.0   281   19   W49736   Protein polymer ad
     28   376.5 37.0   287   19   W49737   Protein polymer ad
     29   376.5 37.0   877   19   R80335   Protein polymeric
     30   376.5 37.0   877   19   W49724   Protein polymer ad
     31   371.5 36.5  2257    9   P82961   SELP3 protein comp
     32   371.5 36.5  2257   14   R41012   SELP3 multimeric p
     33   371.5 36.5  2257   18   W26347   SELP3 synthetic el
     34   371.5 36.5  2257   19   W53523   Amino acid sequenc
     35   371.5 36.5  2257   21   Y78282   SELP3 amino acid s
     36   368.5 36.2   378   19   W09219   SELP0K polymer. S
     37   368.5 36.2   378   19   W53545   Amino acid sequenc
     38   368.5 36.2   378   21   Y51889   Plasmid pPT0375 pr
     39   368.5 36.2  1002   17   W09218   SELP0K polymer.  S
     40   368.5 36.2  1002   19   W53544   Amino acid sequenc
     41   368.5 36.2  1002   21   Y51888   Plasmid pPT0364 pr
     42   366.5 36.0  2055    9   P82960   SEP12 protein comp
     43   366.5 36.0  2055   14   R41011   SELP2 multimeric p
     44   366.5 36.0  2055   18   W26346   SELP2 synthetic el
     45   366.5 36.0  2055   19   W53522   Amino acid sequenc


                                    ALIGNMENTS

RESULT      1
W46316
ID   W46316 standard; protein; 201 AA.
XX
AC   W46316;
XX
DT   23-JUL-1998  (first entry)
XX
DE   Non-natural polypeptide MFU-2 sequence.
XX
KW   MFU-1; minimal functional unit; elastin; human; fibrous protein;
KW   beta-sheet; coating; wound dressing; MFU
XX
OS   Homo sapiens.
XX
PN   W09805685-A2.
XX
PD   12-FEB-1998.
XX
PF   07-AUG-1997;  97WO-CA00560.
XX
PR   07-AUG-1997;  97US-0911364.
PR   07-AUG-1996;  96US-0023552.
XX
PA   (HOSP-) HOSPITAL FOR SICK CHILDREN.
PA   (PROT-) PROTEIN SPECIALTIES LTD.
XX
PI   Keeley FW,  Rothstein A,  Rothstein SJ;
XX
DR   WPI; 1998-145551/13.
XX
PT   New non-natural polypeptide with multiple beta-sheet, beta-turn
PT   structures - particularly based on human elastin, useful for coating
```

FT   prostheses, as wound dressings, etc., allows ingrowth of cells
XX
PS   Claim 9; Fig 4C; English.
XX
CC   This represents the non natural minimal functional unit (MFU)-2
CC   polypeptide which comprises tandem repeats of a portion of the MFU-1
CC   polypeptide of the invention. The MFU is a polypeptide that has at least
CC   3 beta-sheet/ beta-turn structures, but is not a naturally occurring
CC   fibrous protein. Each beta-sheet structure has 3-7 (preferably 5-7) amino
CC   acids and the MFU polypeptide may include at least 1 amino acid that can
CC   take part in crosslinking. The polypeptide can also be derived from the
CC   sequences of animal elastin, lamprin and spider silk protein. The MFU
CC   polypeptides are self-aligning peptides having the same primary structure
CC   as part of a natural fibrous protein. They are used to coat prostheses
CC   made of animal or synthetic material or metal, particularly for use as
CC   blood vessel or heart valve replacements, wound or burn dressings, or
CC   stents. They can be used in cosmetic, elastic or high-tensile strength
CC   materials, e.g. ropes or parachute cord. Prostheses based on the MFU
CC   allow penetration of endothelial cells, so become permanent, living,
CC   tissue replacements. The MFU polypeptides have better biocompatibility
CC   than known elastin-based materials. They are well defined, homogeneous
CC   material and are easier to manipulate and produce than full-length
CC   elastins. They are non-thrombogenic and non-immunogenic. Materials can
CC   be made from 2 or more different MFU polypeptides to allow properties
CC   to be tailored for particular applications, e.g. combining the high
CC   extensibility of elastin and the high tensile strength of spider silk
CC   protein.
XX
SQ   Sequence   201 AA;

Query Match   100.0%;   Score 1017;   DB 19;   Length 201;
Best Local Similarity   100.0%;   Pred. No. 6.2e-68;
Matches   201;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy      1 PPGFGVGGIPGVAPGVGPVGGVPGVGPVGGVPGVGISPEAQAAAAAKAAKYGVGTPAAAAAK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 fpgfgvgvggipgvagvpgvgpvggvpgvgpvggvpgvgispeaqaaaaakaakygvgtpaaaaak 60

Qy     61 AAAKAAQFGLVPGVGPVGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPAIGPPE 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 aaakaaqfglvpgvgpvgvgvapgvgvapgvgvapgvgvapgvgvapaigppe 120

Qy    121 AQAAAAAKAAKYGVGTPAAAAAAKAAAAAKAAQFGLVPGVGPVGVGVAPGVGVAPCVGLAPG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 aqaaaaakaakygvgtpaaaaaakaaaaakaaqfglvpgvgpvgvgvapgvgvapcvglapg 180

Qy    181 VGVAPGVGVAPGVGVAPAIGP 201
          |||||||||||||||||||||
Db    181 vgvapgvgvapgvgvapaigp 201
```

CC   The invention relates to a derivative or variant of human tropoelastin
CC   (hTE) having elastin-like and/or macromolecule (specifically
CC   glycosaminoglycan (GAG))-binding properties. Cells containing vectors
CC   comprising the nucleic acids encoding the variants or derivatives are
CC   used to produce the proteins recombinantly. The tropoelastin derivatives
CC   or hybrid proteins containing the derivatives are useful in medical,
CC   pharmaceutical, veterinary and cosmetic applications, e.g. as anti-
CC   wrinkle or hand lotions, also as surgical implants, foods and industrial
CC   products. The hybrid protein have controllable GAG-binding properties,
CC   depending on presence or absence of a specific fragment, designated
CC   peptide 26A, from hTE. The present sequence represents the synthetic
CC   human tropoelastin variant SHELdelta26A.
XX
SQ   Sequence   698 AA;

Query Match   69.8%;   Score 709.5;   DB 20;   Length 698;
Best Local Similarity   53.8%;   Pred. No. 6.2e-45;
Matches   164;   Conservative   6;   Mismatches   26;   Indels   109;   Gaps   6;

```
Qy      1 PPGFGVGVGGIPGVAPGVGPVGGVPGVGPVGGVPGVGISPEAQAAAAAKAAKYGVGTPAAAAAK 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    378 fpgfgvgvggipgvagvpgvgpvggvpgvgpvggvpgvgispeaqaaaaakaakygvgtpaaaaaK 437

Qy     61 AAAKAAQFGLVPGVGPVGVAPGVGVAPGVGVAPGVGVAPGVGVAPAIGPPE 120
          ||||||||||||||||||||||||||||| ||
Db    438 aaakaaqfglvpgvgpvgvapgvgvapgvglapgvgvapgvgvapgvgvapgvgvapgigpgg 497

Qy    121 AQAA----------------------------------------------- 124
          ||
Db    498 vaaaaksaakvaakaqlraaaglgagipglgvgvgvpglgvgegvpgegvpgfgav 557

Qy    125 ----AAAKAAKYGVGTP--------------------AAAAKAAKAAQFGLVP 155
          ||||||||| |||||||
Db    558 ggalaaakaaakygaavrpgvlgglgaigavggvgipgvgvgagpaaaaakaaakaaqfglgvg 617

Qy    156 GVGVAPGYGV---APGVGVAPGV-------------GLAPGVGVAP--GVGVAPG 192
          |: |:| ||||: ||:: | ||
Db    618 aaglg-glgvggigvpgvvgglggippaaaakaaakygaagiggvlggagiqfpigggvaarpg 676

Qy    193 VGVAP 197
          |:|:|
Db    677 fglsp 681
```

KW   metastasis; blood clotting.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO200004043-A1.
XX
PD   27-JAN-2000.
XX
PF   19-JUL-1999;   99WO-AU00580.
XX
PR   17-JUL-1998;   98AU-0004723.
XX
PA   (UNSI) UNIV SYDNEY.
XX
PI   Weiss AS;
XX
     WPI: 2000-182399/15.
     N-PSDB: 261144.
XX
PT   New derivatives of tropoelastin useful, e.g. for inducing chemotaxis
PT   and cell growth -
XX
PS   Disclosure; Page 110-112; 136pp; English.
XX
CC   The present sequence represents a human reduced tropoelastin derivative,
CC   designated SHEL-delta-26a. The sequence is produced by removing
CC   exon 26a of SHEL (SHEL not defined). The protein is representative of
CC   tropoelastin derivatives of the invention, in which a subsequence has
CC   been mutated so that susceptibility to proteolysis is reduced or
CC   eliminated, or a subsequence has been inserted so that susceptibility
CC   to proteolysis is increased. The derivatives have with reduced
CC   susceptibility, and can be used where the wild-type protein would be
CC   degraded too easily, e.g. in contact with serum or wound exudate. The
CC   tropoelastin derivatives provide competitive inhibition of protease
CC   activity. The tropoelastin derivatives, and other polypeptides containing
CC   tropoelastin derivative-derived protease-susceptibility sites are useful
CC   in human or veterinary medicine, cosmetics (e.g. antiwrinkle or hand
CC   lotions), as bulking agents and for inducing chemotaxis. They are also
CC   useful for proliferation or growth inhibition, particularly of smooth
CC   muscle cells, epithelial or endothelial cells, fibroblasts, osteocytes,
CC   chondrocytes and platelets. Peptidomimetics that mimic the protease
CC   cleavage site in tropoelastin derivatives are competitive inhibitors of
CC   the protease, and are used for protecting against lung damage caused by
CC   elastin, for inhibiting or controlling localized growth of cancers or
CC   metastases, or to limit protease activity that causes blood clotting.
XX
SQ   Sequence   698 AA;

Query Match   69.8%; Score 709.5; DB 21; Length 698;
Best Local Similarity 53.8%; Pred. No. 6.2e-45;
Matches 164; Conservative   6; Mismatches 26; Indels 109; Gaps   6;

Qy   1 FPGFGVGVGGIPGVAGVPGVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGVGTPAAAAAK 60
          !!!
Db   378 fpgfgvgvggipgvagvpgvggvpgvggvpgvggvgtpaaaaak 437

Qy   61 AAAKAAQFGLVPGVGVPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAPAIGPPE 120
          !! !!!
Db   438 aaakaaqfg1vpgvgvpgvgvapgvgvapgvglapgvgvapgvgvapgvgvapg4pgg 497

Qy   121 AQAA-- 124
          !!
Db   498 vaaaaksaakvaakaqlraaaglgasgipglgvgvgvgvpglgvgagvpglgvgagvpgfgav 557

Qy   125 ----AAKAAKYGVGTP-------------AAAAAKAAKAAQFGLVP 155
          !!!!!!!! !!!!!!!!!!!!!!!!
Db   558 pgalaaakaygaavpgvlggigalggvgjipggvvgegpaaaaakaaakaacfglvg 617

Qy   156 GVGVAPGVGV-------APGVGVAPGV----GVGVAPGV 192
          !!!   :!!!   !! | | | !!
Db   618 aaglg-glgvggglgvpgvggglgglgippaaaakaakygaaglggvlggvraairpg 676

Qy   193 VGVAP 197
          |:||
Db   677 fg:sp 681

RESULT   4
Y01303
ID   Y01303 standard; Protein; 660 AA.
XX
AC   Y01303;
XX
DT   07-JUN-1999 (first entry)
XX
DE   Human tropoelastin derivative SHELdeltamodified.
XX
KW   Tropoelastin; hTE; elastin; glycosaminoglycan; GAG-binding; medical;
KW   pharmaceutical; veterinary; cosmetic application; anti-wrinkle; food;
KW   hard lotion; surgical implant; industrial product; human; SHEL.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO9903886-A1.
XX
PD   28-JAN-1999.
XX
PF   17-JUL-1998;   98WO-AU00564.
XX
PR   18-JUL-1997;   97AU-0008117.
XX
PA   (UNSY) UNIV SYDNEY.
XX
PI   Weiss AS;
XX
DR   WPI: 1999-132162/11.
DR   N-PSDB: X27705.
XX
PT   New derivatives of human tropoelastin - with elastin-like or
PT   macromolecular binding properties, useful e.g. as surgical implants
XX
PS   Claim 7; Fig 3; 82pp; English.
XX
CC   The invention relates to a derivative or variant of human tropoelastin
CC   (hTE) having elastin-like and/or macromolecule (specifically
CC   glycosaminoglycan (GAG))-binding properties. Cells containing vectors
CC   comprising the nucleic acids encoding the variants or derivatives are
CC   used to produce the proteins recombinantly. The tropoelastin derivatives
CC   or hybrid proteins containing the derivatives are useful in medical,
CC   pharmaceutical, veterinary and cosmetic applications, e.g. as anti-
CC   wrinkle or hand lotions, also as surgical implants, foods and industrial
CC   products. The hybrid protein have controllable GAG-binding properties,
CC   depending on presence or absence of a specific fragment, designated
CC   peptide 26A, from hTE. The present sequence represents a human
CC   tropoelastin derivative SHELdeltamodified.
XX
SQ   Sequence   660 AA;

Query Match   69.7%; Score 709; DB 20; Length 660;
Best Local Similarity 54.6%; Pred. No. 6.4e-45;
Matches 166; Conservative   6; Mismatches 24; Indels 108; Gaps   7;

Qy   1 FPGFGVGVGGIPGVAGVPGVGGVPGVGGVPGVGISPRAQAAAAAKAAKYGVGTPAAAAAK 60
          !!
Db   341 fpgfgvgvggipgvagvpsvgvvpgvggvpgvgvgvgispeaqaaaaakaakyyvgtpaaaaak 400

Qy   61 AAAKAAQFGLVPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAPAIGPPE 120
          !!! !!!
Db   401 aaakaaqfglvpgvgvapgvggvgvapglapgvgvapgvgvapgigpgg 460

Qy   121 AQAA-- 124
          !!
Db   461 ...

```
Db  461 vaaaaksaakvaakaqlraaaglgaqipglgvgvgvpglgvgagvpglgvgagvpgfgav 520
Qy  125 ---AAKKAAXY------------GVGTP---------AAAAKAPAKRAAQPGLVPG 156
          |||||||                |||         ||||||||||||||||
Db  521 pgalaaakkaakygavpsgrvlgglgalggvgvipggvvgagpaaaaaakaaakaacfglvga 580
Qy  157 VGYAPGVGV---APGVGVAPGV-------------GLAPGVGVAP--GVGVAPGV 193
          |:||      |:||           ||          | ||  ||
Db  581 aglg-glgvgglgvgvpgvglgsippaaaaakaakygaaglgvgaaglgfplggvaarpgf 639
Qy  194 GVAP 197
          |::|
Db  640 glsp 643
```

RESULT 5
B08630
```
ID  B08630 standard; peptide; 712 AA.
XX
AC  B08630;
XX
DT  20-DEC-2000 (first entry)
XX
Amino acid sequence of a human elastin polypeptide.
XX
KW  Tropoelastin; elastin; elastic fibre; smooth muscle cell proliferation;
KW  smooth muscle cell differentiation; smooth muscle cell migration;
KW  smooth muscle cell function; atherosclerosis; restenosis; aneurysm;
KW  vascular bypass graft stenosis; transplant arteriopathy; dissection;
KW  SVAS; hypertension; transplant arteriopathy.
XX
OS  Homo sapiens.
XX
PN  WO20005068-A2.
XX
PD  31-AUG-2000.
XX
PF  28-FEB-2000; 2000WO-US02526.
XX
PR  26-FEB-1999;  99US-0258217.
XX
PA  (UTAH ) UNIV UTAH RES FOUND.
XX
PI  Keating MT, Li DY;
XX
DR  WPI; 2000-533134/48.
XX
PT  Elastin based compositions useful for treating atherosclerosis,
PT  restenosis, vascular bypass graft stenosis, transplant arteriopathy,
PT  aneurysm, dissection SVAS and/or hypertension -
XX
PS  Example 3; Page 46; 79pp; English.
XX
CC  The present sequence represents a human elastin. Peptides derived
CC  from elastin are used in compositions of the invention. The
CC  specification describes elastin based compositions that are potent
CC  regulators of smooth muscle cell proliferation, differentiation and
CC  migration in vivo. The elastin-based compositions comprise at least one
CC  elastic fibre, elastins, tropoelastins (or fragments of them) which have
CC  biological activities comprising: inhibiting the proliferation of smooth
CC  muscle cells in vivo; stimulating the differentiation of smooth muscle
CC  cell in vivo; and regulating the migration of smooth muscle cells in
CC  vivo. The compositions may be used for the prophylaxis or treatment
CC  of a disorder characterized by diminished capacity to regulate smooth
CC  muscle cell function such as atherosclerosis, restenosis, vascular
CC  bypass graft stenosis, transplant arteriopathy, aneurysm and/or
CC  dissection. Disorders which may be treated also include SVAS (undefined),
CC  hypertension, and transplant arteriopathy.
XX
SQ  Sequence   712 AA;
```

Query Match   69.6%; Score 707.5; DB 21; Length 712;

Best Local Similarity 55.7%; Pred. No. 8.8e-45;
Matches 166; Conservative 6; Mismatches 19; Indels 107; Gaps 8;

```
Qy    1 FPGSGVGVGGIPGVAGVPGVGVPGVGVPGVGVGISPEAQAAAAAKAAAKAAKYGVGTPAAAAAK 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  404 fpgfgvvgvgglpgvagvpgvgvpgvgvpgvvgvispeaqaaaaakaaakyayvgtpaaaaak 463
Qy   61 AAKKAAQP-----GLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAP 114
          |:|||||      |||||||||||||||||||||||||||||||||||||||||||
Db  464 aaakaaqfallnlaglvpgvgvapgvgvapgvgvapgvglapgvgvapgvgvapgvgvvap 523
Qy  115 AIGPPEAQAA-------------------------------- 124
          ||| |||
Db  524 gigpggvaaaaaksaakvaakaqlraaaglgagipglgvgvgvpglgvgagvpglgvgagv 583
Qy  125 ---------AAAKAAKYGVGTP------------------------AAAAAKAAAKRAA 149
           |||||||                                 |||||||||
Db  584 pgfgavpgalaaakaakygaavpgrvlgglgalggvgvipggvvgagpaaaaaakaaakaa 643
Qy  150 QFGLVPGVGVAPGVGVAPGVGLAPGVGVAP-----------GVGVAPGVGVAP 197
          |:||| |: ||| |:|| |:|| |:||             || |||:|
Db  644 qfglvgaagig-glgvg-glgv-pgvg---glggippaaaaakaakygvaaarpgfglsp 695
```

RESULT 6
B08631
```
ID  B08631 standard; peptide; 730 AA.
XX
AC  B08631;
XX
DT  20-DEC-2000 (first entry)
XX
Fusion protein comprising human elastin and c-myc.
XX
KW  Tropoelastin; elastin; elastic fibre; smooth muscle cell proliferation;
KW  smooth muscle cell differentiation; smooth muscle cell migration;
KW  smooth muscle cell function; atherosclerosis; restenosis; aneurysm;
KW  vascular bypass graft stenosis; transplant arteriopathy; dissection;
KW  SVAS; hypertension; transplant arteriopathy.
XX
OS  Synthetic.
OS  Homo sapiens.
OS  Unidentified.
XX
PN  WO20005068-A2.
XX
PD  31-AUG-2000.
XX
PF  28-FEB-2000; 2000WO-US02526.
XX
PR  26-FEB-1999;  99US-0258217.
XX
PA  (UTAH ) UNIV UTAH RES FOUND.
XX
PI  Keating MT, Li DY;
XX
DR  WPI; 2000-533134/48.
XX
PT  Elastin based compositions useful for treating atherosclerosis,
PT  restenosis, vascular bypass graft stenosis, transplant arteriopathy,
PT  aneurysm, dissection SVAS and/or hypertension -
XX
PS  Example 3; Page 48; 79pp; English.
XX
CC  The present sequence represents a fusion protein, comprising human
CC  elastin and c-myc, preceded by a His tag. The protein is used in
CC  compositions of the invention. The specification describes elastin
CC  based compositions that are potent regulators of smooth muscle cell
CC  proliferation, differentiation and migration in vivo. The elastin-based
CC  compositions comprise at least one elastic fibre, elastins, tropoelastins
CC  (or fragments of them) which have biological activities comprising:
CC  inhibiting the proliferation of smooth muscle cells in vivo; stimulating
CC  the differentiation of smooth muscle cell in vivo; and regulating the
```

```
CC   migration of smooth muscle cells in vivo. The compositions may be used
CC   for the prophylaxis or treatment of a disorder characterized by
CC   diminished capacity to regulate smooth muscle cell function such as
CC   atherosclerosis, restenosis, vascular bypass graft stenosis, transplant
CC   arteriopathy, aneurysm and/or dissection. Disorders which may be treated
CC   also include SVAS (undefined), hypertension, and transplant arteriopathy.
XX
SQ   Sequence   730 AA;

Query Match             69.6%;  Score 707.5; DB 21;  Length 730;
Best Local Similarity   55.7%;  Pred. No. 9e-45;
Matches 166;  Conservative    6;  Mismatches 19;  Indels 107;  Gaps     8;

Qy    1 FPGFGVGVGGIPGVAGVGVGVGVGVGVGVGVGVISPEAQAAAAKAAKYGVGVTPAAAAAK 60
         |||||||  |||||| | ||||||||||| ||||||||||||||||||||||||||||||
Db  413 fpgfgvgvgg1pgvagvgvgvgvgvgvgvvgvvvgvvgvispeaqaaaaakaakygvgvtpaaaaak 472

Qy   61 ------GLVPGVGVAPGVGVAPGVGVLAPGVGVAPGVGVAP 114
               |||||||||||||||||||||||||||||||||
Db  473 aaakaaqfglnllaglVpgvgvgvapgvgvgvlapgvgvgvrgvgvgvgvvap 532

Qy  115 AIGPPEAQAA------ 124
         ||  |||
Db  533 gigpggvaaaksaakvaakaqlraaaglgaglgvgvgvgtpg1gvgvgvtpg1gvgvagv 592

Qy  125 ------AAAKAAKYGVGTP------ 149
               ||||||||||
Db  593 psfgavpgallaaakaaakygaavpgvlgg1galggvgvgipggvvgagvpgpaaaakaakaakaa 652

Qy  150 QEGLVPGVGVAPGVGVAPGVGVAPGVGVLAPGVGVAP--------GVGVAPGVGVAP 197
         |||| |: |:|| |: || |||| |||||                |||   |||||:|
Db  653 qfglvgaag1g-glgvg-g1gv-pgvg--giggippaaakaakygvaarpgfgisp 704
```

```
RESULT  7
W46315
ID    W46315 standard; protein; 730 AA.
XX
AC    W46315;
XX
DT    23-JUL-1998  (first entry)
XX
DE    Human elastin minimal functional non-natural polypeptide MFU-1 sequence.
XX
KW    MFU-1; minimal functional unit; elastin; human; fibrous protein;
KW    beta-sheet; coating; wound dressing.
XX
OS    Homo sapiens.
XX
FH    Key             Location/Qualifiers
FT    Protein         374..499
FT                    /note= "MFU-1 polypeptide"
XX
PN    WO9805685-A2.
XX
PD    12-FEB-1998.
XX
PF    07-AUG-1997;    97WO-CA00560.
XX
PR    07-AUG-1997;    97US-0911364.
PR    07-AUG-1996;    96US-023552.
XX
PA    (HOSP-) HOSPITAL FOR SICK CHILDREN.
PA    (PROF-) PROTEIN SPECIALTIES LTD.
XX
PI    Keeley FW, Rothstein A, Rothstein SJ;
XX
DR    WPI; 1998-145551/13.
XX
PT    New non-natural polypeptide with multiple beta-sheet, beta-turn
PT    structures - particularly based on human elastin, useful for coating
PT    prostheses, as wound dressings, etc., allows ingrowth of cells
```

```
XX
PS    Claim 5; Fig 1B; 39pp; English.
XX
CC    This represents the human elastin sequence containing the minimal
CC    functional unit (MFU)-1 polypeptide of the invention. This MFU-1 is a
CC    polypeptide that has at least 3 beta-sheet/ beta-turn structures, but
CC    is not a naturally occurring fibrous protein. Each beta-sheet structure
CC    has 3-7 (preferably 5-7) amino acids and the MFU polypeptide may include
CC    at least one amino acid that can take part in crosslinking. The polypeptide
CC    can also be derived from the sequences of animal elastin, lamprin and
CC    spider silk protein. The MFU polypeptides are self-aligning peptides
CC    having the same primary structure as part of a natural fibrous protein.
CC    They are used to coat prostheses made of animal or synthetic material or
CC    metal, particularly for use as blood vessel or heart valve replacements,
CC    wound or burn dressings, or stents. They can be used in cosmetic,
CC    elastic or high-tensile strength materials, e.g. ropes or parachute cord.
CC    prostheses based on the MFU allow penetration of endothelial cells, so
CC    become permanent, living, tissue replacements. The MFU polypeptides have
CC    better biocompatibility than known elastin-based materials. They are well
CC    defined, homogeneous material and are easier to manipulate and produce
CC    than full-length elastins. They are non-thrombogenic and non-immunogenic.
CC    Materials can be made from 2 or more different MFU polypeptides to allow
CC    properties to be tailored for particular applications, e.g. combining the
CC    high extensibility of elastin and the high tensile strength of spider
CC    silk protein.
XX
SQ   Sequence   730 AA;

Query Match             68.8%;  Score 700; DB 19;  Length 730;
Best Local Similarity   76.1%;  Pred. No. 3.2e-44;
Matches 156;  Conservative    7;  Mismatches 24;  Indels 18;  Gaps     8;

Qy    1 FPGFGVGVGGIPGVAGVPGVGVGVGVGVGVGVGVGVISPEAQAAAAAKAAKAAKYGVGTPAAAAAK 60
         |||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db  378 fpgfgvgvvgg1pgvagvpgvagvpgvsgvvgvvgvyvyvyvispeaqaaaaakaakygvgtpaaaaak 437

Qy   61 AAAKAAQFGLVPGVGVAPGVGVAPGVGVLAPGVGVAPGVGVAPGVGVAPAIGPPE 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  438 aaakaaqfglvrpgvgvgvgvapgvgvgvgvlapgvgvgvapgvgvtapgigp-- 495

Qy  121 AQAAAAAKAAKYGVGTPAAAAAKAAAKAAKAAQPGL--VPGVGVAPGVGVAPGV-G 176
         |||||:|        |  | |||||||||:||   ||| :||:|| ||||||||||||| ||
Db  496 ggvaaaaksa------akvaakaqlraaa-g1gagipglgv--gvgv-pglgvgagvpg 544

Qy  177 LAPGVGVAPGVGVAPGVGVAPAIGP 201
         ||| |||:|  | ||||||:| :||
Db  545 lgvgagvr-pgfgagadegvrrislsp 568
```

```
RESULT  8
Y69068
ID    Y69068 standard; Protein; 731 AA.
XX
AC    Y69068;
XX
DT    30-MAY-2000  (first entry)
XX
DE    Amino acid sequence of a human tropoelastin splice form.
XX
KW    Tropoelastin; derivative; proteolysis; protease; antiwrinkle;
KW    hand lotion; bulking agent; chemotaxis; proliferation;
KW    growth inhibition; peptidomimetic; lung damage; elastin; cancer;
KW    metastasis; blood clotting.
XX
OS    Homo sapiens.
XX
FH    Key             Location/Qualifiers
FT    Cleavage-site   515..516
FT    Cleavage-site   564..565
FT    Cleavage-site   441..442
FT    Cleavage-site   503..504
XX
```

PN   WO200004043-A1.
XX
PD   27-JAN-2000.
XX
PF   19-JUL-1999;   99WO-AU00580.
XX
PR   17-JUL-1998;   98AU-0004723.
XX
PA   (UNSY) UNIV SYDNEY.
XX
PI   Weiss AS;
XX
DR   WPI; 2000-182399/16.
DR   N-PSDB; Z61146.
XX
PT   New derivatives of tropoelastin useful, e.g. for inducing chemotaxis and cell growth -
XX
PS   Disclosure; Page 107-109; 136pp; English.
XX
CC   The present sequence represents a human tropoelastin splice form. The specification describes tropoelastin derivatives, in which a subsequence has been mutated so that susceptibility to proteolysis is reduced or eliminated, or a subsequence has been inserted so that susceptibility to proteolysis is increased. The derivatives have with reduced susceptibility, and can be used where the wild-type protein would be degraded too easily, e.g. in contact with serum or wound exudate. The tropoelastin derivatives provide competitive inhibition of protease activity. The tropoelastin derivatives, and other polypeptides containing tropoelastin derivative-derived protease-susceptibility sites, are useful in human or veterinary medicine, cosmetics (e.g. antiwrinkle or hand lotions), as bulking agents and for inducing chemotaxis. They are also useful for proliferation or growth inhibition, particularly of smooth muscle cells, epithelial or endothelial cells, fibroblasts, osteocytes, chondrocytes and platelets. Peptidomimetics that mimic the protease cleavage site in tropoelastin derivatives are competitive inhibitors of the protease, and are used for protecting against lung damage caused by elastin, for inhibiting or controlling localized growth of cancers or metastases, or to limit protease activity that causes blood clotting.
XX
SQ   Sequence 731 AAs;

Query Match   68.8%; Score 700; DB 21; Length 731;
Best Local Similarity   76.1%; Pred. No. 3.2e-44;
Matches 156; Conservative   7; Mismatches 24; Indels 18; Gaps 8;

Qy   1 FPGFGVGVGGIPGVAGVPGVGGVPGVGGVPGVGISPEAQAAAAKAAKYVGTPAAAAAK 60
          ||||||||||:|||||||| :|||| ||||||||||||||||||||||||||||||||
Db 378 fpgfgvgvgggipgvagvpsvgvgvpgvggvpgvgvgispeaqaaaaakaakygvgtpaaaaak 437

Qy   61 AAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPAIGPPE 120
          ||||||||||||||||||||||||||||||||||||||||||||||| ||
Db 438 aaakaaqfglvpgvgvapgvgvapgvgvapgvgvapgvgvapgigp-- 495

Qy 121 AQAAAAAKAAKYGVGTPAAAAKAAKAAQFGL---VPGVGVAPGVGVAPGV-G 176
          :|||:| || |||:|| || :|:|| || |||:||||||||||| ||
Db 496 ggvaaaaksa------akvaakaqlraaa-glgagipglgv--gvgv-pglgvgagvpg 544

Qy 177 LAPGVGVAPGVGVAPGVGVAPAIGP 201
          | ||||||:|| || ::|:|
Db 545 lgvgagv-pgfgagadegvrrslsp 568

RESULT   9
R56653
ID   R56653 standard; Protein; 733 AA.
XX
AC   R56653;
XX
DT   22-MAR-1995 (first entry)
XX
DE   Synthetic human tropoelastin (SHEL).
XX
KW   Tropoelastin; pharmaceutical; surgical dressing.
XX
OS   Synthetic.
XX
PN   WO9414958-A.
XX
PD   07-JUL-1994.
XX
PF   16-DEC-1993;   93WO-AU00655.
XX
PR   22-DEC-1992;   92AU-0006520.
PR   28-JUN-1993;   93AU-0009661.
XX
PA   (UNSY) UNIV SYDNEY.
XX
PI   Martin SL, Weiss AS;
XX
DR   WPI; 1994-263533/32.
DR   N-PSDB; Q70941.
XX
PT   Synthetic polynucleotide(s) - encode recombinant tropoelastins and variants
XX
PS   Disclosure; Page 30; 77pp; English.
XX
CC   Human synthetic tropoelastin is susceptible to hydrolytic breakdown of the crosslinks. Such material may be useful in e.g. surgical applications, where the gradual loss of material over time is intended.
XX
SQ   Sequence 733 AAs;

Query Match   68.8%; Score 700; DB 15; Length 733;
Best Local Similarity   76.1%; Pred. No. 3.2e-44;
Matches 156; Conservative   7; Mismatches 24; Indels 18; Gaps 8;

Qy   1 FPGFGVGVGGIPGVAGVPGVGGVPGVGGVPGVGISPEAQAAAAKAAKYVGTPAAAAAK 60
          ||||||||||:|||||||| :|||| ||||||||||||||||||||||||||||||||
Db 380 fpgfgvgvgggipgvagvpsvgvgvpgvggvpgvgvgispeaqaaaaakaakygvgtpaaaaak 439

Qy   61 AAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPAIGPPE 120
          ||||||||||||||||||||||||||||||||||||||||||||||| ||
Db 440 aaakaaqfglvpgvgvapgvgvapgvgvapgvgvapgvgvapgigp-- 497

Qy 121 AQAAAAAKAAKYGVGTPAAAAKAAKAAQFGL---VPGVGVAPGVGVAPGV-G 176
          :|||:| || |||:|| || :|:|| || |||:||||||||||| ||
Db 498 ggvaaaaksa------akvaakaqlraaa-glgagipglgv--gvgv-pglgvgagvpg 546

Qy 177 LAPGVGVAPGVGVAPGVGVAPAIGP 201
          | ||||||:|| || ::|:|
Db 547 lgvgagv-pgfgagadegvrrslsp 570

RESULT   10
Y01301
ID   Y01301 standard; Protein; 733 AA.
XX
AC   Y01301;
XX
DT   07-JUN-1999 (first entry)
XX
DE   Amino acid sequence of synthetic human tropoelastin SHEL.
XX
KW   Tropoelastin; hTE; elastin; glycosaminoglycan; GAG-binding; medical; pharmaceutical; veterinary; cosmetic application; anti-wrinkle; food; hand lotion; surgical implant; industrial product; human; SHEL.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9903886-A1.

XX
PD   28-JAN-1999.
XX
PF   17-JUL-1998;   98WO-AU00564.
XX
PR   18-JUL-1997;   97AU-0008117.
XX
PA   (UNSY) UNIV SYDNEY.
XX
PI   Weiss AS;
XX
DR   WPI; 1999-132162/11.
DR   N-PSDB; X27704.
XX
PT   New derivatives of human tropoelastin - with elastin-like or
PT   macromolecular binding properties, useful e.g. as surgical implants
XX
     Disclosure: Fig 1; 82pp; English.
XX
CC   The invention relates to a derivative or variant of human tropoelastin
CC   (hTE) having elastin-like and/or macromolecule (specifically
CC   glycosaminoglycan (GAG)-binding properties. Cells containing vectors
CC   comprising the nucleic acids encoding the variants or derivatives are
CC   used to produce the proteins recombinantly. The tropoelastin derivatives
CC   or hybrid proteins containing the derivatives are useful in medical,
CC   pharmaceutical, veterinary and cosmetic applications, e.g. as anti-
CC   wrinkle or hand lotions, also as surgical implants, foods and industrial
CC   products. The hybrid protein have controllable GAG-binding properties.
CC   depending on presence or absence of a specific fragment, designated
CC   peptide 26A, from hTE. The present sequence represents the amino acid
CC   sequence of the synthetic human tropoelastin SHEL.
XX
SQ   Sequence 733 AA;

Query Match   58.8%;   Score 700; DB 20; Length 733;
Best Local Similarity 76.1%;   Pred. No. 3.2e-44;
Matches 156; Conservative   7; Mismatches 24; Indels 18; Gaps 8;

Qy   1 FPGFGVGVGGIPGVAGVPGVGGVPGVGVPGVGISPEAQAAAAAKAAKYGVGTPAAAAAK 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   380 fpgfgvgvggipgvagvpgvgsvpgvgvpgvgvpgvgvispeaqaaaakaakygvgtpaaaaak 439

Qy   61 AARKAAQPGLVPGVGVPGVGVAPGVGVAPGVGVSLAPGVGVAPGVGVARAIGPPE 120
          ||||| : : || :||:::   :|:
Db   440 aakaaqfgivpgvgvapgvgvapgvgvapgvglapgvgvapgvgvapgvgvapgigp-- 497

Qy   121 AQRAAAAAAKYGVGTPAAAKAAKAAKAAQPGL--VPGVGVAPGVGVAPGVGVAPGV-G 176
          ||||| : :|| :||:| :|| | | :||:|| ||:|| |||||
Db   498 ggvaaaksa-------akvaaaqlraaa-glgagipglgv--gvgv-pglgvgagvpg 546

Qy   177 LARPGVGVAPGVGVAPAIGP 201
          |||||||| | : :|
Db   547 lgvgagv-pgfgagadegvrrislsp 570

RESULT 11
ID   Y69071 standard; Protein; 571 AA.
XX
AC   Y69071;
XX
DT   30-MAY-2000 (first entry)
XX
DE   Amino acid sequence of a human tropoelastin derivative.
XX
KW   Tropoelastin; derivative; proteolysis; protease; antiwrinkle;
KW   hand lotion; bulking agent; chemotaxis; proliferation;
KW   growth inhibition; peptidomimetic; lung damage; elastin; cancer;
KW   metastasis; blood clotting.
XX
OS   Homo sapiens.
XX

PN   WO20000004043-A1.
XX
PD   27-JAN-2000.
XX
PF   19-JUL-1999;   99WO-AU00580.
XX
PR   17-JUL-1998;   98AU-0004723.
XX
PA   (UNSY) UNIV SYDNEY.
XX
PI   Weiss AS;
XX
DR   WPI; 2000-182399/16.
XX
PT   New derivatives of tropoelastin useful, e.g. for inducing chemotaxis
PT   and cell growth -
XX
PS   Disclosure: Page 115-117; 136pp; English.
XX
CC   The present sequence represents a human tropoelastin derivative, which
CC   is representative of tropoelastin derivatives of the invention. In the
CC   tropoelastin derivatives of the invention a subsequence has been
CC   mutated so that susceptibility to proteolysis is reduced or eliminated,
CC   or a subsequence has been inserted so that susceptibility to proteolysis
CC   is increased. The derivatives have with reduced susceptibility, and can
CC   be used where the wild-type protein would be degraded too easily,
CC   e.g. in contact with serum or wound exudate. The tropoelastin
CC   derivatives provide competitive inhibition of protease activity. The
CC   tropoelastin derivatives, and other polypeptides containing tropoelastin
CC   derivative-derived protease-susceptibility sites, are useful in human
CC   or veterinary medicine, cosmetics (e.g. antiwrinkle or hand lotions), as
CC   bulking agents and for inducing chemotaxis. They are also useful for
CC   proliferation or growth inhibition, particularly of smooth muscle cells,
CC   epithelial or endothelial cells, fibroblasts, osteocytes, chondrocytes,
CC   and platelets. Peptidomimetics that mimic the protease cleavage site in
CC   tropoelastin derivatives are competitive inhibitors of the protease, and
CC   are used for protecting against lung damage caused by elastin, for
CC   inhibiting or controlling localized growth of cancers or metastases, or
CC   to limit protease activity that causes blood clotting.
XX
SQ   Sequence 571 AA;

Query Match   67.2%;   Score 683.5; DB 21; Length 571;
Best Local Similarity 79.4%;   Pred. No. 4.1e-43;
Matches 150; Conservative   6; Mismatches 14; Indels 19; Gaps 7;

Qy   1 FPGFGVGVGGIPGVAGVPGVGGVPGVGVPGVGISPEAQAAAAAKAAKYGVGTPAAAAAK 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   378 fpgfgvgvggipgvagvpgvgsvpgvgvpgvgvpgvgvispeaqaaaakaakygvgtpaaaaak 437

Qy   61 AARKAAQPGLVPGVGVPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAPAIGPPE 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   438 aakaaqfgivpgvgvapgvgvapgvgvapgvglapgvgvapgvgvapgvgvapgigp-- 495

Qy   121 AQRAAAAAKAAKYGVGTPAAAAARAAKAAQPGL---VPGVGVAPGVGVAPGVGL 177
          ||||| :   : | || : | |||| |||| ||:|| |||| ||
Db   496 ggvaaaaksa------akvaaaqlraaa-glgagipglgv--gvgv-pglgvgagv-- 542

Qy   178 APGVGVAPG 186
          ||:|| |
Db   543 -pglgvgag 550

RESULT 12
Y69135
ID   Y69135 standard; Protein; 515 AA.
XX
AC   Y69135;
XX
DT   30-MAY-2000 (first entry)
XX
DE   Amino acid sequence of a human tropoelastin derivative.

XX
KW Tropoelastin; derivative; proteolysis; protease; antiwrinkle;
KW hand lotion; bulking agent; chemotaxis; proliferation;
KW growth inhibition; peptidomimetic; lung damage; elastin; cancer;
KW metastasis; blood clotting.
XX
OS Homo sapiens.
XX
PN WO2000004043-A1.
XX
PD 27-JAN-2000.
XX
PF 19-JUL-1999; 99WO-AU00580.
XX
PR 17-JUL-1998; 98AU-0004723.
XX
PA (UNSY) UNIV SYDNEY.
XX
PI Weiss AS;
XX
DR WPI; 2000-182399/16.
XX

New derivatives of tropoelastin useful, e.g. for inducing chemotaxis
and cell growth -

XX
PS Disclosure: Page 131-133; 136pp; English.
XX
CC The present sequence represents a human tropoelastin derivative, which
CC is representative of tropoelastin derivatives of the invention. In the
CC tropoelastin derivatives of the invention a subsequence has been
CC mutated so that susceptibility to proteolysis is reduced or eliminated,
CC or a subsequence has been inserted so that susceptibility to proteolysis
CC is increased. The derivatives have with reduced susceptibility, and can
CC be used where the wild-type protein would be degraded too easily,
CC e.g. in contact with serum or wound exudate. The tropoelastin
CC derivatives provide competitive inhibition of protease activity. The
CC tropoelastin derivatives, and other polypeptides containing tropoelastin
CC derivative-derived protease-susceptibility sites, are useful in human
CC or veterinary medicine, cosmetics (e.g. antiwrinkle or hand lotions), as
CC bulking agents and for inducing chemotaxis. They are also useful for
CC proliferation or growth inhibition, particularly of smooth muscle cells,
CC epithelial or endothelial cells, fibroblasts, osteocytes, chondrocytes
CC and platelets. Peptidomimetics that mimic the protease cleavage site in
CC tropoelastin derivatives are competitive inhibitors of the protease, and
CC are used for protecting against lung damage caused by elastin, for
CC inhibiting or controlling localized growth of cancers or metastases, or
CC to limit protease activity that causes blood clotting.
XX
SQ Sequence 515 AA;

Query Match   65.9%; Score 670; DB 21; Length 515;
Best Local Similarity 59.3%; Pred. No. 3.7e-42;
Matches 156; Conservative   9; Mismatches   9; Indels 26; Gaps   9;

Qy   4 FGVGVGGI-PGV-AGVPGV-GGVPGVGGVPGVGI-SPEAQAAAAAKAAXY--------- 49
            | | |: | ::| |||||| | :|:|| | | :|||||||||
Db 240 fgagaagvlpgvgsagvpgvTpgvvpgaipgigvgiiagvgtpaaaaaaaakaakygaaagivpg 299

Qy   50 --------------GVGTPAA----------------AAAKAAAKAAQPGL 70
                          |||| :|||||||||||:|
Db 300 gpgfpgvgvgvgpgvpgvvpgagipgvvpgagipgaavpgvvspeaaakaakyga 359

Qy   71 VPGVGVA--------------PGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPAIGP----- 118
            |||| ||| ||| |: :|||||||:|
Db 360 rpgvgvgipTygvgagpfpgfgv--gvggipgvagvTpgvggvpgvggvpsvggygv-gis---- 412

Qy 119 PEAQAAAAAKAAKYGVGTPAAQAAKAAKAAAKAAQFGLVPGVGVAPGVGVAPGVGVGLA 178
            ||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db 413 peeqaaaaakaakyavgtpaaaaaaaakaaakaaaqkaaqTgivpgvvpgvgvapgvgla 472

Qy 179 PGVGVAPGVGVAPGVGVAPAIGP 201
            |||||||||||||| ||||| ||

Db 473 pgvgvapgvgvapgvgvapgigp 495

RESULT 13
Y31682
ID Y31682 standard; Protein; 745 AA.
XX
AC Y31682;
XX
DT 22-NOV-1999 (first entry)
XX
DE Alanine-containing elastomer, polymer IV.
XX
KW Elastomer; bioelastomer; polymer IV; tissue augmentation;
KW tissue restoration; tissue reconstruction; tissue repair;
KW implant; fibronectin cell attachment site.
XX
OS Synthetic.
XX
PN WO9943271-A1.
XX
PD 02-SEP-1999.
XX
PF 25-FEB-1999; 99WO-US04440.
XX
PR 29-MAY-1998; 98US-0087155.
PR 27-FEB-1998; 98US-0076297.
XX
PA (BIOE-) BIOELASTICS RES LTD.
XX
PI Glazer PA; Parker TM; Urry DW;
XX
DR WPI; 1999-540487/45.
XX
PT Augmentation or restoration of mammalian tissue by injecting
PT solution of peptide polymer, used for soft or hard tissue
PT reconstruction, especially of intervertebral disks
XX
PS Claim 73; Page 106-109; 133pp; English.
XX
CC The present sequence represents novel alanine-containing polymer IV
CC that has bioelastic properties. The invention provides a method of
CC tissue augmentation by injecting a polymer comprising repeating
CC peptide monomeric units selected from nonapeptide, pentapeptide and
CC tetrapeptide monomeric units, where the monomeric units form a
CC series of beta-turns separated by dynamic bridging segments. The
CC polymer has an inverse temperature transition value (Tt) that is
CC less than the tissue temperature, and is injected in water solution
CC at coacervate concentration. The polymer is injected at periurethral
CC or subdermal sites (for treatment of urinary incontinence or for
CC cosmetic purposes), or into hard or soft tissue, e.g. for repair of
CC traumatic injury. A specific application is restoration of
CC intervertebral discs. Polymer IV has a Tt value of 28 deg.
XX
SQ Sequence 745 AA;

Query Match   52.3%; Score 531.5; DB 20; Length 745;
Best Local Similarity 45.2%; Pred. No. 7.1e-33;
Matches 126; Conservative 46; Mismatches 48; Indels 99; Gaps   9;

Qy   2 PGFGV-GVG----GIPGVAGVPGVG----GVPGVG----GVPGVGISPEAQAAAAAKAAK 48
            ||:| | | :|| |:|| | |||| |||:
Db 138 pgvgvpgvvpgvgvggvpgvgvgv-gtpgvygvpgvgvypgvgvTpgvTpgvgvgvgvgv- 183

Qy   49 YGVGTPAAAAAAKAAKAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAP 108
            || | :||||||||||||||:|| ||||||||||||||:||||:|||||||||||
Db 184 -gvgvpgvgvpgv--gvapgvgvapgvgvapgvgvapgvgvapgvvapgvgvap 240

Qy 109 GVGVAPAIGPP--- 119
            ||||| :||
Db 241 gvgvapgvvpgvvpgvgvgvgvgvgvgvgvpgvgvgvpgvgvgvgvpgvgvgvpgvgvgvgv 300

```
Qy   120 ------------EAQAAAAKAKAKYGVGTPAAAAAKAAAKAAQFGLVPGVGVGVAP 161
                                     ||| |  |  |::  ||||||
Db   301 pgvgvpgvgvgvgvgvgvgvgvgvgvgvgvgvpgvv--gvapgvgvap 358

Qy   162 GVGVAPGVGVGLAPGVGVAPGVGVAPGVGVAPAIG 200
             |||||||||||||||||||||||||||| :|
Db   359 gvgvapgvgvapgvgvapgvgvapgvgvapgvgvapgvgvg 397

RESULT 14
Y31681
ID   Y31681 standard; protein; 148 AA.
XX
AC   Y31681;
XX
DT   22-NOV-1999 (first entry).
XX
    Alanine-containing elastomer monomeric unit.
XX
KW   Elastomer; bioelastomer; polymer; tissue augmentation;
KW   tissue restoration; tissue reconstruction; tissue repair;
KW   implant; fibronectin cell attachment site.
XX
OS   Synthetic.
XX
PN   WO9943271-A1.
XX
PD   02-SEP-1999.
XX
PF   26-FEB-1999;      99WO-US04440.
XX
PR   29-MAY-1998;      98US-0087155.
PR   27-FEB-1998;      98US-0076297.
XX
PA   (BIOE-) BIOELASTICS RES LTD.
XX
PI   Glazer PA; Parker TM; Urry DW;
XX
DR   WPI; 1999-540487/45.
DR   N-PSDB; X87852.
XX
PT   Augmentation or restoration of mammalian tissue by injecting
PT   solution of peptide polymer, used for soft or hard tissue
PT   reconstruction, especially of intervertebral disks
XX
    Claim 72; Fig 6; 133pp; English.
CC
CC   The present sequence represents the monomeric unit of a novel
CC   alanine-containing polymer (see Y31682) that has bioelastic
CC   properties. The invention provides a method of tissue
CC   augmentation by injecting a polymer comprising repeating peptide
CC   monomeric units selected from nonapeptide, pentapeptide and
CC   tetrapeptide monomeric units, where the monomeric units form a
CC   series of beta-turns separated by dynamic bridging segments. The
CC   polymer has an inverse temperature transition value that is less
CC   than the tissue temperature, and is injected in water solution at
CC   coacervate concentration. The polymer is injected at periurethral
CC   or subdermal sites (for treatment of urinary incontinence or for
CC   cosmetic purposes), or into hard or soft tissue, e.g. for repair of
CC   traumatic injury. A specific application is restoration of
CC   intervertebral discs.
XX
SQ   Sequence   148 AA;

Query Match     44.8%;  Score 455.5;  DB 20;  Length 148;
Best Local Similarity   56.4%;  Pred. No. 6.2e-27;
Matches  114;  Conservative    6;  Mismatches   15;  Indels   67;  Caps

Qy     2 PGPGVGGIPGVAGVPGVG----GVPGVG---GVPGVGISPEPAQAALAAAKAKYGVGT 53
           ||| ||  |:||   |||||   |||||:   |||||           |||
Db     5 pgvgvpgvgvgv-gvpgvvgvgvgvgvgvgvgvgvp------------gvgv 49
```

```
Qy    54 PAAAAAKAKAKAAQFGLVPGVGVGVAPGVGVAPGVGVLAPGVGVAPGVGVGVAPGVGVA 113
           |: ||||||||||||||||||||||||||||::|||||||||||||
Db    50 pgv----------gvapgvgvapgvgvapgvgvapgvgvapgvgvgvapgvgvgva 97

Qy   114 PAIGPPEAQAAAAKAKYGVGTPAAAAAKAAAKAAAQFGLVPGVGVGVAPGVGVAPGVGVAP 173
           | |:|       |||     |||  |||||||||||||||||||||  ||||| |
Db    98 pgvgvp------------gvgvpgvg-------------vpgvgv-pgvgvp-pgvgv-p 128

Qy   174 GVGLAPGVGVAPGVGVAPGVGV 195
           ||: |||| ||||| ||||||
Db   129 gvgv-pgvgv-pgvgv-pgvgv 147

RESULT 15
P82484
ID   P82484 standard; protein; 294 AA.
XX
AC   P82484;
XX
DT   31-OCT-1990 (first entry)
XX
DE   Tropoelastin.
XX
KW   Chicken; tropoelastin; platelet aggregation; arteriosclerosis.
XX
OS   Gallus domesticus.
XX
FH   Key          Location/Qualifiers
FT   Region       3..32
FT                /label=repeat_region
FT                /rpt_type=tandem
FT                /rpt_unit= RU1:VPGVG
FT   Region       53..54
FT                /label=repeat_unit
FT                /rpt_type=direct
FT                /rpt_unit= RU2:AAAAKAAAKAAK
FT   Region       92..96
FT                /label=RU1
FT   Region       99..103
FT                /label=RU1
FT   Region       119..123
FT                /label=RU1
FT   Region       125..137
FT                /label=RU2
FT   Region       173..182
FT                /label=rpt_region
FT                /rpt_type=tandem
FT                /rpt_unit=RU1 (x2)
XX
PN   JP63014694-A.
XX
DT   21-JAN-1988.
XX
PF   08-JUL-1986;      86JP-0158655.
XX
PR   08-JUL-1986;      86JP-0158655.
XX
PA   (KAOS ) KAO CORP.
PA   (SHIS ) SHISEIDO KK.
XX
DR   WPI; 1988-059829/09.
DR   N-PSDB; N82138.
XX
PT   cDNA fragment encoding tropoelastin - has inhibiting activity towards
PT   platelet aggregation and is useful for prophylaxis of arteriosclerosis.
XX
PS   Claim 1; Fig 1; 5pp; Japanese.
XX
CC   The cDNA was sequenced from a clone prepd. from mRNA isolated from
CC   chicken aorta tissue. It can be used to produce tropoelastin which
CC   is a precursor of elastin which does not have desmosine, isodesmosine
CC   and lysynoylleucine residues and isn't cross-linked.
XX
```

SQ   Sequence   294 AA:

Query Match   43.9%;   Score 446.5;   DB 9;   Length 294;
Best Local Similarity   54.1%;   Pred. No. 5.2e-26;
Matches   120;   Conservative   11;   Mismatches   44;   Indels   47;   Caps   16;

```
Qy      2 PGPGVGVGGIPGVAGPGVPGVG-----GVPGVG-------GVPGVGI------SPEAQAAAAAKA 46
            || || | :| || ||||||       |||||         |||||:         | : | ||||||

Db      4 pgvgvpgvpgv-gvpgvgvpgvgvgvpgvgvpgvgvPpglvpgagpaaaakaaaka 62

Qy     47 AKYGVGTPAAAAAKAAAKAAAQFGLVPGV-GVAPGV-GVAPGV-GVAPGV-GLAPGV-GVA 101
            |||: |                |:|||  || :|||| |: ||| || | ||| ||  |:|

Db     63 akygag-------------glapgvvgglapgvvggiapgvvggilvPgvvggla 106

Qy    102 PGV-GVAPGVGVAPAIGPPEAQAAAAKAAAKYGVGTPAAAAAKAAAKAAQFGLVPGV-GV 159
            ||| |:|| ||  |:| | | i | | |||||||||  |  |          |||| ||

Db    107 pgvgglapavgavpgvvggpaaaakaaakYgagvvggvvpgavag---vpgvpgv 162

Qy    160 APGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAPAIGP 201
            ||||  |  : ||||: ||||| |:|| | |:

De    163 tpgvggvp--sivpgvgv-pgvgvlpgagi-pgvgvqpgqrkp 200
```

Search completed: April 24, 2001, 16:38:15
Job time: 415 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:　April 24, 2001, 16:41:50 ; Search time 74.56 Seconds
(without alignments)
185.264 Million cell updates/sec

Title:　US-09-340-736-2
Perfect score:　1017
Sequence:　1 FPGFGVGVGGIPGVAGVPGV.......GVAPGVGVAPGVGVAPAIGP 201

Scoring table:　BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:　198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters:　198801

Minimum DB seq length:- 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :　PIR_67:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	662.5	65.1	792	2	EAHU	elastin precursor,
2	545.5	53.6	770	2	S59623	tropoelastin, she
3	485.7	47.7	747	2	S39623	elastin precursor,
4	474	46.6	784	1	A26601	elastin precursor
5	412	40.5	864	1	EAR7	elastin precursor
6	379	37.3	860	2	EAMS	elastin precursor
7	346.5	34.1	907	2	A45560	sporozoite surface
8	249.5	24.5	76	2	I45885	elastin - bovine (
9	237	23.3	1334	2	T50568	probable multi-dom
10	230	22.6	1489	2	D70807	hypothetical glyci
11	223	21.9	1901	2	F70806	hypothetical glyci
12	216.5	21.3	416	1	SKXLAG	dermal gland prote
13	214.5	21.1	445	2	A27263	50K spicule matrix
14	209.5	20.6	641	2	Q03831	nuclear antigen EB
15	205.5	20.2	1660	2	A70869	hypothetical glyci
16	198.5	19.5	713	2	URMS	period clock prote
17	195	19.2	419	2	G70602	hypothetical prote
18	192.5	18.9	627	2	A44112	hypothetical glyci
19	192	18.9	2639	2	T31328	spidroin 2, dragli
20	188	18.5	743	2	S23779	fibroin - Chinese
21	187.5	18.4	1306	2	A70934	collagen alpha 1(V
22	186.5	18.3	744	2	S15435	collagen alpha 1(V
23	186.5	18.3	767	2	E70895	hypothetical glyci
24	185	18.2	416	2	T32458	hypothetical prote
25	185	18.2	744	2	A34246	collagen alpha 1(V
26	185	18.2	744	2	S23298	collagen alpha 1(V
27	184.5	18.1	1763	2	S16356	collagen alpha 2(I
28	184	18.1	498	2	C70720	hypothetical glyci
29	183.5	18.0	413	2	E70651	probable PE protei
30	182.5	17.9	278	2	S44796	F09G8.6 protein -
31	180.5	17.7	408	2	S57483	glycin-rich protei
32	180.5	17.7	1381	2	E70806	hypothetical glyci
33	180.5	17.7	1669	1	CGHC4B	collagen alpha 1(I
34	180	17.7	460	2	T33110	hypothetical prote
35	180	17.7	1373	2	A43291	collagen alpha 2(I
36	179.5	17.6	384	1	A26099	glycine-rich cell
37	178	17.5	584	2	G70804	hypothetical glyci
38	178	17.5	615	2	H70589	hypothetical glyci
39	177.5	17.5	490	2	T09084	phosphatidylinosit
40	177	17.4	783	2	T35389	probable serine-th
41	177	17.4	1691	1	S22917	collagen alpha 5(I
42	176.5	17.4	308	2	T37286	collagen 40 - Caen
43	176	17.3	778	2	F70963	hypothetical glyci
44	175.5	17.3	882	2	F70812	hypothetical glyci
45	175	17.2	923	2	E70820	hypothetical glyci

ALIGNMENTS

RESULT 1
EAHU
elastin precursor, long splice form - human
N;Alternate names: tropoelastin
C;Species: Homo sapiens (man)
C;Date: 22-Jun-1990 #sequence_revision 26-Jul-1996 #text_change 22-Jun-1999
C;Accession: A32707; A33705; A30524; A53891
R;Indik, Z.; Yeh, H.; Ornstein-Goldstein, N.; Sheppard, P.; Anderson, N.; Rosenbloom,
Proc. Natl. Acad. Sci. U.S.A. 84, 5680-5684, 1987
A;Title: Alternative splicing of human elastin mRNA indicated by sequence analysis of
A;Reference number: A32707; MUID:87289658
A;Accession: A32707
A;Molecule type: mRNA
A;Residues: 1-500,507-792 <IND>
A;Cross-references: GB:M16983; GB:J02948
R;Bashir, M.M.; Indik, Z.; Yeh, H.; Ornstein-Goldstein, N.; Rosenbloom, J.C.; Abrams,
J. Biol. Chem. 264, 8887-8891, 1989
A;Title: Characterization of the complete human elastin gene. Delineation of unusual
A;Reference number: A33705; MUID:89255358
A;Accession: A33705
A;Residues: 1-27 <BAS>
A;Cross-references: GB:J04821; NID:g182052; PIDN:AAA52379.1; PID:g553276
R;Fazio, M.J.; Olsen, D.R.; Kauh, E.A.; Baldwin, C.T.; Indik, Z.; Ornstein-Goldstein,
J. Invest. Dermatol. 91, 458-464, 1988
A;Title: Cloning of full-length elastin cDNAs from a human skin fibroblast recombinan
A;Reference number: A30524; MUID:89009960
A;Accession: A30524
A;Molecule type: mRNA
A;Residues: 1-453,483-617,651-792 <FA2>
A;Cross-references: EMBL:M36860; NID:g182061; PIDN:AAA52382.1; PID:g182062
A;Note: this sequence represents a composite of several splice forms
R;Fazio, M.J.; Olsen, D.R.; Kuivaniemi, H.; Chu, M.L.; Davidson, J.M.; Rosenbloom, J.
Lab. Invest. 58, 270-277, 1988
A;Title: Isolation and characterization of human elastin cDNAs, and age-associated va
A;Reference number: A53891; MUID:88156138
A;Accession: A53891
A;Molecule type: mRNA
A;Residues: 164-453,483-500,507-617,651-792 <FA2>
A;Cross-references: GB:M24782; NID:g182063; PIDN:AAA53190.1; PID:g182064
C;Comment: The term tropoelastin refers to a soluble precursor form of the extracellu
ine oxidase activity.
C;Genetics:
A;Gene: GDB:ELN
A;Cross-references: GDB:119107; OMIM:130160
A;Map position: 7q11.23-7q11.23
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-792/Product: elastin #status predicted <MAT>
F;782-787/Disulfide bonds: #status predicted

Left column

```
Query Match       65.1%;   Score 562.5;  DB 1;  Length 792;
Best Local Similarity  65.0%;  Pred. No. 2.5e-33;
Matches 156;  Conservative  53;  Indels  24;  Mismatches  7;  Gaps  10;

Qy    1  FPGFGVGVGGIPGVAGVPGVPGVGPGVGPGVPGVGISPEAQAAAAKAAKY----------  49
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  404  FPGFGVGVGGIPGVAGVPGVPGVGGVPGVGPGVGISPEAQAAAARAAKYGRAAGAGVLGGL  463

Qy   50  --------------------------GVGTPAAAAAKAAKAAQF------GIVPGVGVAPGVGVAPG   85
                                   ||||||||||||||        ||||||||||||||
Db  464  VPGPQAANPGVPGTGGVPGVGGTPAAAAAKAAKAAQFALLNLAGLVPGVGVAPGVGVAPG  523

Qy   86  VGVAPGVGLAPGVGVAPGVGVAPGVGVAPAIGPPEAQAAAAKAAKYGVGTPAAAAAKAA  145
         |||||||||||||||||||||||||||||||| ||     |||
Db  524  VGVAPGVGLAPGVGVAPGVGVAPGVGVAPGIGP--------GGVAAAAAKSA------AKYAAKAQ  574

Qy  146  AKAAQFGL---VPGVGVAPGVGVAPGVGVAPGV-GLAPGVGVAPGVGVAPGVGVAPAIGP  201
         :|| :||    :||:|    |||:|| |:||    ||| :|| ||   ||  ::|
Db  575  LRAAA-GLGAGIPGLGV--GVGV-PGILGVGAGVPGLGVGAGV-PGFGAGAGDEGVRNSLSP  629

Query Match       53.6%;   Score 545.5;  DB 2;  Length 770;
Best Local Similarity  50.9%;  Pred. No. 2.6e-26;
Matches 139;  Conservative  11;  Indels  24;  Mismatches  11;  Gaps  14;

Qy    5  GVGVGGIPGVAGVPGV-GGVPGVGGVPGVGISPE--AQAAAAAKAAKYG----------v--VG   52
         ||||| |  |:||:|| | ::|||  |:| | :|:  | ||||||||||||
Db  289  GVGVGG----AGIPGAGAIPGIGGIAGAG-APDAAAAAAKAAKYGAAGVFGPGVVG  343

Qy   53  TPAAAAAKAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGV  112
         || |    |||||||||||||||| ||||  |||| |||| :||| |||| ||||
Db  344  VPGAGVPGVG--------VPGVGV-PGVGV-PGVGV-PGVGI-PGVGY  389

Qy  113  A--------------------------------PAIGPPEAQAAAAAKAAKY-  132
         |                                ||||||||||| |||||||
Db  390  PGAVSPAAAAKAAAKAAKFGAAGGVGVGGIPTPGVPGGPFGPGGDAAAQAAAARAAAKIG  449

Qy  133  --------------------------GVGTPAAAAAKAAKAAQFGLVPGVGVAPGVG  164
                                    ||  |||||||||||| |||||||||||||
Db  450  AGGVGLGGLVPGAGAIPGVPGVGGVPGVGGVPGVGRPAAAAAKAAKAAQFGLCPGVGVAPGVG  509

Qy  165  VAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAP  197
         :|||||||||||:||||||||||:|| :|:
Db  510  VAPGVGVAPGVGVPGVGVSVAPGVGVAPGIGLGP  542
```

Right column

```
RESULT  3
EABO
elastin precursor, splice form a - bovine
N:Alternate names: tropoelastin
N:Contains: elastin precursor, splice form b; elastin precursor, splice form c
C:Species: Bos primigenius taurus (cattle)
C:Date: 08-Jun-1989 #sequence_revision 26-Jul-1996 #text_change 22-Jun-1999
C:Accession: A31865; A26728; B26728; C26728; A22343; I45886
R:Yeh, H.; Anderson, N.; Ornstein-Goldstein, N.; Bashir, M.M.; Rosenbloom, J.C.; Abra
Biochemistry 28, 2365-2370, 1989
A:Title: Structure of the bovine elastin gene and S1 nuclease analysis of alternative
A:Reference number: A31865; MUID:89274159
A:Accession: A31865
A:Molecule type: DNA
A:Residues: 1-27 <YEH>
R:Raju, K.; Anwar, R.A.
J. Biol. Chem. 262, 5755-5762, 1987
A:Title: Primary structures of bovine elastin a, b, and c deduced from the sequences
A:Reference number: A9264O; MUID:87194772
A:Accession: A26728
A:Molecule type: mRNA
A:Cross-references: GB:J02855; NID:g340504; PIDN:AAA30776.1; PID:g552339
A:Accession: B26728
A:Molecule type: mRNA
A:Residues: 1,'RS',4-11,'E',13-636,'V',638-747 <RA>
A:Cross-references: GB:J02717; NID:g163019; PIDN:AAA30503.1; PID:g163020
A:Accession: C26728
A:Molecule type: mRNA
A:Residues: 1,'RS',4-11,'E',13-225,240-636,'V',638-747 <RA2>
A:Cross-references: GB:K03505; NID:g163025; PIDN:AAA30505.1; PID:g163026
A:Accession: C26728
A:Molecule type: mRNA
A:Residues: 1,'RS',4-11,'E',13-225,260-636,'V',638-747 <RA3>
A:Cross-references: GB:K03506; NID:g163027; PIDN:AAA30506.1; PID:g163028
R:Cicila, G.; May, M.; Ornstein-Goldstein, N.; Indik, Z.; Yeh, H.; Morrow, S.; Yeh, H.S.; Rose
Biochemistry 24, 3075-3080, 1985
A:Title: Structure of the 3' portion of the bovine elastin gene.
A:Reference number: A22343; MUID:85280426
A:Accession: A22343
A:Molecule type: DNA
A:Residues: 613-747 <CIC>
A:Cross-references: GB:M20415
R:Rosenbloom, J.
Lab. Invest. 51, 605-623, 1984
A:Title: Biology of disease: Elastin: Relation of protein and gene structure to disea
A:Reference number: I45885; MUID:85052254
A:Accession: I45886
A:Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 6/8-683,685-747 <ROS>
A:Cross-references: GB:M91898; NID:g163015; PIDN:AAA96417.1; PID:g163018
R:Brown, P.L.; Mecham, L.; Tisdale, C.; Mecham, R.P.
Biochem. Biophys. Res. Commun. 186, 549-555, 1992
A:Title: The cysteine residues in the carboxy terminal domain of tropoelastin form an
A:Reference number: A58621 MUID:92337651
A:Contents: annotation, disulfide bonds
C:Comment: the term tropoelastin refers to a soluble precursor form of the extracellu
ine oxidase activity.
C:Genetics:
A:Introns: 634/3; 653/3; 676/3; 689/3; 707/3; 716/3; 733/3
A:Note: the list of introns is incomplete
C:Superfamily: elastin
C:Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F:1-747/Product: elastin precursor, splice form a #status predicted <EPA>
F:1-225,260-747/Product: elastin precursor, splice form c #status predicted <EPC>
F:1-225,240-747/Product: elastin precursor, splice form b #status predicted <EPB>
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-747/Product: elastin #status predicted <MAT>
F:105,109,252,271,275,324,327,400,404,407,445,448,489,493,544,552,606,609,645,649
F:737-742/Disulfide bonds: #status experimental

Query Match       47.7%;   Score 485.5;  DB 1;  Length 747;
Best Local Similarity  47.8%;  Pred. No. 1e-22;
Matches 129;  Conservative  13;  Mismatches  37;  Indels  91;  Gaps  14;
```

Below left column

```
RESULT  2
S23
tropoelastin - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 23-Aug-1996 #sequence_revision 13-Mar-1997 #text_change 24-Oct-1997
C:Accession: S59623; A24758
R:Mauch, J.C.; Sandberg, L.B.; Roos, P.J.; Jimenez, F.; Christiano, A.M.; Deak, S.B.; Bo
Matrix Biol. 14, 635-641, 1994
A:Title: Extensive alternate exon usage at the 5' end of the sheep tropoelastin gene.
A:Reference number: S59623
A:Accession: S59623
A:Status: Preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-770 <MAU>
R:Yoon, K.; Davidson, J.M.; Boyd, C.; May, M.; LuValle, P.; Ornstein-Goldstein, N.; Smit
Arch. Biochem. Biophys. 241, 684-691, 1985
A:Title: Analysis of the 3' region of the sheep elastin gene.
A:Reference number: A24758; MUID:85305763
A:Accession: A24758
A:Molecule type: mRNA
A:Residues: 555-669,671-716,732-770 <YOO>
C:Superfamily: elastin
C:Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F:760-765/Disulfide bonds: #status predicted
```

```
Qy      5 GVGVGIPGVAVPGVGVPGVGVPGVGVPGVGISPE--AQAAAAKAAKY------GVGTPA    55
          |||||| | :||:||:| : |:||:||:| |||:|| :|  |.|||||||||
Db    286 GVGVGG-PGIPGAP--GAIPSIGGIAGVG-APDAAAAAAAAAKAAKFGRAGGLPGVGVPG   341

Qy     56 AAAAAKAAKAQFGLVPGVVAPGVGVGVGLAPGVGVAPGVGLAPGVGVAPGVGVGVTAPA   215
          | || |||||:||| || |||:||| || ||| ||||| |||||| ||||| :| :|
Db    342 VGVPGVGVPGVG---VPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVGVPGA       392

Qy    116 IGP--------------------------------------PEAQLAAAAAKY---   132
          |||                                     | |||||||||| |||
Db    393 iSPAAAAAKAAAKAAAKFGARGAVGIGGIPTFGLGPGGPFGIGDAAAAAPAAAAAKAAAKIGAG   452

Qy    133 ---------------------GVGTPAAAAAKAAAKAAQFGLVPGVGVAPGVGVA       166
                               |:||| ||||||||||||:||||:|||||||||
Db    453 GVGALGGVVPGAGAIPGLPGVGVPGVGVAPGVGVAPGVGVV                   512

Qy    167 PGVGVAPGVGLAPGVGVAPGVGVAPGVGVA 196
          ||||| || :|| ||:| : || :|| /
Db    513 PGVGVVPGVGVAPGIGLGPGGVIGAAVPAA 542
```

RESULT 4
A26601
elastin precursor - chicken (fragment)
N;Alternate names: tropoelastin
C;Species: Gallus gallus (chicken)
C;Date: 05-Oct-1988 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
C;Accession: A26601; A30795; A27264
R;Bressan, G.M.; Argos, P.; Stanley, K.K.
Biochemistry 26, 1497-1503, 1987
A;Title: Repeating structure of chick tropoelastin revealed by complementary DNA cloning.
A;Reference number: A26601; MUID:87242320
A;Accession: A26601
A;Molecule type: mRNA
A;Residues: 1-1212,237-524,535-784 <BR2>
A;Cross-references: GB:M15889; NID:g212803; PIDN:AAA49108.1; PID:g212804
R;Baule, V.J.; Foster, J.A.
Biochem. Biophys. Res. Commun. 154, 1054-1060, 1988
A;Title: Multiple chick tropoelastin mRNAs.
A;Reference number: A30795; MUID:88309083
A;Accession: A30795
A;Molecule type: mRNA
A;Residues: 85-784 <BAD>
A;Cross-references: GB:M21880; NID:g212741; PIDN:AAA49082.1; PID:g212742
R;Kimitsu, T.; Tajima, S.; Nishikawa, T.; Tajima, M.; Fukasawa, T.
J. Biochem. Biophys. 256, 455-461, 1987
A;Title: Sequence analysis of elastin cDNA from chick aorta and tissue-specific transcr
A;Reference number: A27264; MUID:87297334
A;Accession: A27264
A;Molecule type: mRNA
A;Residues: 491-569,'G',571-604,'A',605-643,'A',645-687,'R',689-700,'R',702-784 <TOK>
A;Cross-references: GB:M18633; NID:g211742; PIDN:AAA48761.1; PID:g211743
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;1-24/Domain: signal sequence (fragment) #status predicted <SIG>
F;25-212,237-524,535-784/Product: elastin #status predicted <MAT>
F;773-779/Disulfide bonds: #status predicted

```
Query Match      46.6%;   Score 474;   DB 2;  Length 784;
Best Local Similarity  56.4%;  Pred. No. 5.2e-22;
Matches 132;  Conservative 12;  Mismatches 48;  Indels 42;  Gaps 19;

Qy      2 PGF---GVGVGGIPGVAGVPGVGVPGVGVPGVGVPGVGISPEAQAAAAAAAKYGVGT---PA  55
          ||| | |||||| | |||:| |||:||| |||:||| :||| |  |.|||||      |
Db    387 APGVPGVGIGGVPGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGPAA         446

Qy     56 AAAKAAKAAQFGL--VPGVGV--APGVGVAPGVGVAPGVGLAPGVGVAPGVGVGV----  106
          ||||||||||||    |||||   |||||||||||:|||||||||||||| |||:|||
Db    447 AAAKAAKAAAPGAGRVPGVGVGNVPGVGV-PGVGV-PGVGV-PGVGVPGVG-PVGVG  502

Qy    107 APGVGV----APAIGPPEAQAAAAAKAAKYGVGTPA-------AAAKAAAKAAQF---G 152
          ||||||     |   |  || |||||||||||||||       |||||||||||   |
Db    503 VPGVGVPGVGVGVP-----GVGVGGVGVPGVGVGVVPGAGPAAAKAAAKAAAKYGAGG  559
```

```
Qy    153 LVPGAP-GVAPGV-GVAPGV-GVAPGV-GLAPGV-GVAPGV-GVAPGVGVAPAIG     200
          ||||| | | ||| | |||| | |||| | |||| | |||| :| ||| |||:|
Db    560 LAPGVGGLAPAVGGLAPGVGVGGLVPGVGGGLAPGVGGLAPGVGVGGLAPGVGVPGVG     613
```

RESULT 5
clastin precursor - rat
N;Alternate names: tropoelastin
C;Species: Rattus norvegicus (Norway rat)
C;Date: 11-Jan-1991 #sequence_revision 16-Aug-1996 #text_change 22-Jun-1999
C;Accession: A36106; A30878; A36523; S02173; I54172; I68505
R;Pierce, R.A.; Deak, S.B.; Stolle, C.A.; Boyd, C.D.
Biochemistry 29, 9677-9683, 1990
A;Title: Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.
A;Reference number: A36106; MUID:91104868
A;Accession: A36106
A;Molecule type: mRNA
A;Residues: 1-864 <PIE>
A;Cross-references: GB:M60647; GB:J05292; NID:g207444; PIDN:AAA42269.1; PID:g207445
R;Deak, S.B.; Pierce, R.A.; Belsky, S.A.; Riley, D.J.; Boyd, C.D.
J. Biol. Chem. 263, 13504-13507, 1988
A;Title: Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.
A;Reference number: A30878; MUID:88330068
A;Accession: A30878
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 781-864 <DEA>
A;Cross-references: GB:J04035; NID:g207442; PIDN:AAA42268.1; PID:g207443
R;Franzblau, C.; Pratt, C.A.; Faris, B.; Colannino, N.M.; Offner, G.D.; Mcgaylel Jr.,
J. Biol. Chem. 264, 15115-15119, 1989
A;Title: Role of tropoelastin fragmentation in elastogenesis in rat smooth muscle cel
A;Reference number: A36523; MUID:89355327
A;Accession: A36523
A;Molecule type: protein
A;Residues: 22-31 <FRR>
R;Rich, C.B.; Foster, J.A.
Arch. Biochem. Biophys. 268, 551-558, 1989
A;Title: Characterization of rat heart tropoelastin.
A;Reference number: S02173; MUID:89117149
A;Accession: S02173
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: IP',369-545,548-764,770-864 <RIC>
R;Pierce, R.A.; Alatawi, A.; Deak, S.B.; Boyd, C.D.
Genomics 12, 651-658, 1992
A;Title: Elements of the rat tropoelastin gene associated with alternative splicing.
A;Reference number: I54172; MUID:92241859
A;Accession: I54172
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 264-533 <RES>
A;Cross-references: GB:M86372; NID:g207455; PIDN:AAA42271.1; PID:g554527
A;Accession: I68505
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 558-864 <RE4>
A;Cross-references: GB:M86376; NID:g207459; PIDN:AAA42272.1; PID:g207462
C;Genetics:
A;Introns: 277/1,292/1,308/1,339/1,359/1,419/1,437/1,467/1,484/1,601/1,621/
A;Note: the list of introns may be incomplete
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-864/Product: elastin #status predicted <MAT>
F;854-859/Disulfide bonds: #status predicted
```

```
Query Match      40.5%;   Score 412;   DB 1;  Length 864;
```

```
Best Local Similarity  47.7%;  Pred. No. 3e-18;
Matches 116;  Conservative  12;  Mismatches  43;  Indels  72;  Gaps  18;

Qy      2  PGFGV-GVGGIPGVAGVPGVGPGVGGVPGVG----------ISPEAGAAAAKAAKY-   49
              |||||  ::|::|||||| :|::||  :| |  ||||||||  :|:
Db    379  PGVGIPGVGGIPGVGGIPGVGGIPGVGG-PGIGGPGIVGGPGAVSPAAAAKRAKAAKYG  437

Qy     50  ---GVGTPA------------------------RAAAKAAAKRAAQPGLVPGVGVAPGV--   80
                  |||:|                  |||| ||||||| :|:: |
Db    438  ARGGVGIPPTYGVGAGGPPGVGVGAGGLGGASQRAAAAAAAAKRAAKYG-AGGAGILGGLVP  496

Qy     81  GVAPGV--GVAPGV--GLAPGV--GVAPGVGVAPGVGVAPGVGVAPEAQAAAAKAAKYGV  134
              |:||||  :||:: |  ||||| :| | ||: ||: ||: |||
Db    497  GAVPGALPGAVPGALPGAVPGV---PGTGGVPGAGPP----AAA--------AAA-----  542

Qy    135  GTPAAAAKAAAKRAAQPGLVPGVGVAPGVGVAPGVGVLAPGVGVAPGVGVVAPGVG  194
              ||||||||||||||:: ||| :|||  |:|  || ||| :|: |:| |:|
Db    543  ---AAAAKAAAKRAGQYGLGPGVGGVPG-GV--GVGGLPG-GVGPG-GVT-GIGTGPGTG  593

Qy    195  VAP  197
              : :
Db    594  LVP  596
```

RESULT   6
EAMS
elastin precursor - mouse
N;Alternate names: tropoelastin
C;Species: Mus musculus (house mouse)
C;Date: 18-Aug-1995 #sequence_revision 16-Aug-1996 #text_change 22-Jun-1999
C;Accession: A55721
R;Wydner, K.S.; Sechler, J.L.; Boyd, C.D.; Passmore, H.C.
Genomics 23, 125-131, 1994
A;Title: Use of an intron length polymorphism to localize the tropoelastin gene to mouse
A;Accession: A55721
A;Molecule type: mRNA
A;Residues: 1-860 <WYD>
A;Cross-references: GB:U08210; NID:g473273; PIDN:AAA80155.1; PID:g473274
C;Genetics:
A;Map position: 5
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-860/Product: elastin #status predicted <MAT>
F;850-855/Disulfide bonds: #status predicted

```
Query Match          37.3%;  Score 379;  DB 1;  Length 860;
Best Local Similarity 38.3%;  Pred. No. 2.9e-15;
Matches 128;  Conservative  11;  Mismatches  57;  Indels 138;  Gaps  20;

Qy      2  PGFGVGVGGIPGVA----------GVP-----------------GVGGVPGVGGV   29
              |||| |  |  |||            |||                  | |  | |
Db    243  PGGGGAPSGIPGVGPFGGGQQPGVPLGIYPIKAPKLPGGYGLPYTNGKLPYGVAGAGGKAGY  302

Qy     30  P-GVGISPEAQARAAAAKAAKY----------------------GVGTP-AAA   57
              | |||:|| | |||||||||                         |||| |||
Db    303  PTGTGVGSQA-RAAAKAAKYGAGGAGVLPGVGGGGIPGGAGAIPGIGAGAGTPAAAA  361

Qy     58  AAKAAAKAAQF---GLVP-GVGV-APGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVG  111
              ||||||||:::    |||: | |  | ||||||| |: ||| ||| :|: |:|
Db    362  AAKAAAKAAKYGAGGLVPGGSVRLPGAGI-PGVGGIPGVGGIPGVG-GPGIG-GPGIV  418

Qy    112  VAPAIGPPEAQARAAAAKAAY-----GVGTP----------AARAAKA  144
              ||| |::: |: |:|||||  |       ||||            : |||
Db    419  GGPGAVSPAAAAKAAAKAAKYGAKGGVGIPTYGVGAGGFPGVGVGAGGLGGASPRAAAA  478

Qy    145  AAKRAQP----------GLVPGV-GVAPG-VGVAPGVGVAPG---------------  174
              |||: |          ||||| |:| | | :||:| :| |
Db    479  AAKAAKYGAGGAGGALGGLVPGAVPGALPGAVPVPGAGTPAAAAAAAAKAKAAK  538

Qy    175  VGLAPGVGVAP---GVGVAPG---VGVAPAIGP  201
              : :
```

```
Db    539  AGLGPGVGGVPGGVGVGGIPGVGGVPGGVGVGP  572
              || ||||  |  ||||| |  ||  | |||  ::|
```

RESULT   7
A45560
sporozoite surface antigen SPAG-1 - Theileria annulata
C;Species: Theileria annulata
C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C;Accession: A45560
R;Hall, R.; Hunt, P.D.; Carrington, M.; Simmons, D.; Williamson, S.; Mecham, R.P.; Ta
Mol. Biochem. Parasitol. 53, 105-112, 1992
A;Title: Mimicry of elastin repetitive motifs by Theileria annulata sporozoite surfac
A;Reference number: A45560; MUID:92365719
A;Accession: A45560
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-907 <HAL>
A;Cross-references: GB:M63017; NID:g161884; PID:g161885
A;Note: sequence extracted from NCBI backbone (NCBIN:111148, NCBIP:111150)
C;Keywords: surface antigen

```
Query Match          34.1%;  Score 346.5;  DB 2;  Length 907;
Best Local Similarity 39.1%;  Pred. No. 2.7e-14;
Matches 111;  Conservative  14;  Mismatches  64;  Indels  95;  Gaps  16;

Qy      2  PGFGV-GVGG----GIPGVAGVPGVG----GVPGVGGVPGVGISPEAQAAAAAKAAKYGVG   52
              ||: || ||     :|||:| | ||     :|| ||||  |||::|
Db    179  PGVGVPGVPGVPGVGVPGV-GVPGVGPGVGGVPGVGVGVPGVGVPGVGVPGVGVAPGVGVG  237

Qy     53  T------PAAAAKAAAKAAQPGLVPGVGVAPGVGVA------PGVGLAPGVGVAP  102
              :         |||  | ||| :| | ::      |||  |: :|: || :|:
Db    238  ADSSGLPGSGGLGAGAKAGK-GQQSGLQGPGGVGVVPGVGVAASSSPGKPGVGAGVMP  296

Qy    103  GVGV-----------APGVGVAPA--IGPPEAQ-----------  122
              ||||            ||||||  : ||: |
Db    297  GVGVRAQGGVIIGAPGVAGVPGGKPGQPVSQELELKSDTEINESSSSEBDDDEEEEE  356

Qy    123  --------AAAAAAKAYGVGTPAAAAAKAAAKAA--------QFGLV-----------  154
              |||||:| :|::      :|:  :  | |||
Db    357  ENKSTSSKGAGGKAGK-GQGSVSPGGGSSASQTSPTTTPQSGLASSGSHAQQSPCQQDPAP  415

Qy    155  ---PGVGVAPGVSVAPGVAPGVGVGVLAPGVGVAPGVGVAPGVGVA  196
              |: ::|::|| || || :|  : :| | ||||||  |:
Db    416  SKPSGGGVPGVGV-PGVGV-PGVGV-PGVGVAPGVGVPGVGGA  456
```

RESULT   8
I45885
elastin - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change 13-Aug-1999
C;Accession: I45885
R;Rosenbloom, J.
Lab. Invest. 51, 605-623, 1984
A;Title: Biology of disease: Elastin: Relation of protein and gene structure to disea
A;Reference number: I45885; MUID:85050254
A;Accession: I45885
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-76 <ROS>
A;Cross-references: GB:M31891; NID:g163008; PIDN:AAA96416.1; PID:g552319
C;Genetics:
A;Introns: 20/1; 58/1
C;Superfamily: elastin

```
Query Match          24.5%;  Score 249.5;  DB 2;  Length 76;
Best Local Similarity 60.8%;  Pred. No. 2.4e-09;
Matches  62;  Conservative   2;  Mismatches   9;  Indels  29;  Gaps   4;

Qy     52  GTPAAA--AAKAAAKAAQPGLVPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPG  109
```

```
Db            2 GVPAAAKSAAKAAAK-AQFRLGPGVGVAPGVGVGVGVPGVGVAPGNGLGPG------ 54

Qy          110 VGVAPAIGPPEAQAAAAKAAAKYGVGTPAAAAKAAAKYGVGTPA 151

Db           55 -GV-----------------IGVGTPAAAAKAAAKAAQF 76

RESULT 9
T50568
probable multi-domain regulatory protein [imported] - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T50568
R;Redenbach, M.; Kieser, H.M.; Denapaite, D.; Eichner, A.; Cullum, J.; Kinashi, H.; Hopw
Mol. Microbiol. 21, 77-96, 1995
A;Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb S
reference number: Z20556; MUID:97000351
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1334 <RED>
A;Cross-references: EMBL:AL133220; PIDN:CAB61705.1
A;Experimental source: strain A3(2)
C;Genetics:
A;Note: SCC75A.05c

Query Match            23.3%; Score 237; DB 2; Length 1334;
Best Local Similarity  34.4%; Pred. No. 1.4e-07;
Matches 73; Conservative 5; Mismatches 96; Indels 38; Gaps 6;

Qy           12 PGVAGVPGVGVPGVGVPGVGVGISPEAQAAAAAKAAAKYGVGTPAAAAAKAAAKYAAQFGLV 71

Db          251 FSPTPGRSRFPGWTSGPG----PASGAGAAS------GTDVASGAGAAS-----GPD 294

Qy           72 PGVGVAPGVGVGVPGVAPGVGVAPGVGVAPGVGVAP------GVGVAPAIGP---PEAQ 122

Db          295 PASGPASGPAVAPGSGGPAPGWPAPSTAPGSSTAPPHDTASAADTAPAGPTSAPGTA 354

Qy          123 AAAAKAAKYGVGTPAGPTAGPAGPGTPAVAGTTPAPGTAPAPGPARDTSYAPGTAPVA 414
              ||||  ||  |  ||||||  |  ||| |
Db          355 PAAGTAPAPGTAGPAPGSTPAPGTVPAPGTAPAPGP 446

Qy          170 GVAPGVGLAPGVGVAPGVGVAPGVGVAPGVGVAPAIGP 201
              ||||||| ||  |  ||  ||| ||| ||
Db          415 GTTPAPGTAPAPGSTPAPGTVPAPGTAPAPGP 446

RESULT 10
D70807
hypothetical glycine-rich protein Rv3514 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: D70807
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Skelton, K.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1489 <COL>
A;Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17751.1; PID:g292445
C;Genetics:
A;Gene: Rv3514
C;Superfamily: collagen alpha 1(IV) chain
```

```
Query Match            22.6%; Score 230; DB 2; Length 1489;
Best Local Similarity  35.5%; Pred. No. 4e-07;
Matches 82; Conservative 8; Mismatches 107; Indels 34; Gaps 5;

Qy            3 GFGVCVGGIPGVAGVPGVGVPGVGVPGVGVGISPEAQAAAAAKAAAKYGVGTPAAAAAK 60

Db          340 GDGV-LGGVGGIGGKGGVGGVAGLGGAGGAAGQLFSASGAGNAGVGGAGGQGGDGGAGG 398

Qy           61 AAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGVAPGVGLAPGVVAPGVGVAPGVGVAPAIGPP- 119

Db          399 AGADADQPGATGGTGFAGGAGGAGGGGSSGAGGTNGSGGANGGQGGAGGAGGAGGADNPTG 458

Qy          120 --------EAQAAAAKAAKYG------VGTPAAA--------AAKAAKAA 149

Db          459 IGGTGCGGTGGATGGGGAGGTAGTGGTGGMIGTTGNAGVGGAGGAGGTGGCGGAGGAGGAGG 518

Qy          150 QPGLVPGVGVAPGVVAPGVGVGVGLAPGVGVAPGVGVAPAIG 200

Db          519 QPGATGGTGTAGGAGGAGGAGGSSGAGGTNGSGGAGGTGGCGGAGGAGGAGG 569

RESULT 11
F70806
hypothetical glycine-rich protein Rv3508 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: F70806
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: F70806
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1901 <COL>
A;Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17745.1; PID:g292
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3508
C;Superfamily: collagen alpha 1(IV) chain

Query Match            21.9%; Score 223; DB 2; Length 1901;
Best Local Similarity  32.9%; Pred. No. 1.3e-06;
Matches 77; Conservative 7; Mismatches 114; Indels 36; Gaps 4;

Qy            3 QFGV--GVGGIPGVAGVPGVGVPGVGVPGVGVGISPEAQAAAAAKAAAKYGVGTPAAAAAK 60

Db          340 GDGVAGGVGGTGGKGGVGGVAGLGGAGGAAGQDFSAGGAAGAVGVGGTGGCGGAGGAGAA 399

Qy           61 AAAKAAQFGLVPGVGVAPGVG-------VAPGVVAPGVGVGLAPGVGVAPGVG------- 105

Db          400 GADAFASTGLIGGGTGFAGGAGGVGGQGGNAIAGGINGSGGAGGTGGQGGAGGMGGSGADN 459

Qy          106 ----------VAPGVGVPAIGPPEAQAAAAKAAAKYIGVGTP--AAAAAKAA 146

Db          460 ASGIGADGGAGGTGGNAGAGGAGGAGGTGGCGGVAPGVGLAPGVGVAPGVGVAPAIG 519

Qy          147 KAAQFGLVPGVGVAPGTGGTGGRGGGTGGVAGLGGAGGAAGQLFSAGGAAGGVGVGGTGGCGGAGGAIG 200

Db          520 DATATGATGGTGTGFSGGAGGAGGAGGAGGNTGVGGTNGSGGGGTGGAGGAGGAGGWG 573

RESULT 12
S54XLAG
dermal gland protein APEG precursor - African clawed frog (fragment)
C;Species: Xenopus laevis (African clawed frog)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jun-2000
C;Accession: S07498; A34140
R;Gmachl, M.; Berger, H.; Thalhammer, J.; Kreil, G.
```

FEBS Lett. 260, 145-148, 1990
A;Title: Dermal glands of Xenopus laevis contain a polypeptide with a highly repetitive
A;Reference number: A34140; MUID:90127399
A;Accession: S07498
A;Molecule type: mRNA
A;Residues: 1-416 <OMA>
A;Cross-references: EMBL;X51394; NID:g64547; PIDN:CAA35759.1; PID:g1334643
A;Note: the authors translated the codon TTC for residue 8 as Pro
C;Superfamily: dermal gland protein APEG; trefoil homology
C;Keywords: skin; tandem repeat
F;1-20/Domain: signal sequence (fragment) #status predicted <SIG>
F;21-416/Product: dermal gland protein APEG #status predicted <MAT>
F;23-344/Region: alanine/glutamic acid/glycine/proline-rich repeats
F;349-389/Domain: trefoil homology <TRF>
F;349-375,359-374,369-386/Disulfide bonds: #status predicted

Query Match   21.3%;   Score 216.5; DB 1; Length 416;
Best Local Similarity   36.0%; Pred. No. 9.3e-07;
Matches   77; Conservative   6; Mismatches 102; Indels   29; Gaps   7;

Qy   5 GVGVGGIP----GVGVGPVGVGGVP------GVGGVPGVGISPEAQRAAAAAAKXAKYGVGTP 54
            | ||| ||| |:||:|| |||||
Db   20 GQDAGGEPPAPAEGVAPAPAEGGAPAPAPAEGGAPAPAEGGAPAPAPAEGAPPAPAPADGGAP 79

Qy   55 AAAAAKAAKXAAQFGLVPGVGVAPGVGVAPGVGVLAPGVGVAPG--VGVAPGVGV 112
            | ||| || || ||||| |||
Db   80 APAPAEGGAPAP---APAEGGAPAPAPAEGGAPAPAEGGAPAPAPAEGEAPAPAPAEGE 135

Qy   113 APAIGPPEAQAAAAAKA------AKYGVGIPAAAAAAAAAKKAAQFGLVPGVGVAPGVGVAP 167
            ||:| | :|| :|; |:| | || |
Db   136 APAPAPAEGEAPAPAEGGAPAPAPAEVEAPAPAPAEGEAPAP----APAEGEAPAPAPAP 191

Qy   168 GVGVAPGVGLAPGVGVAPGVGVGVAPAIGP 201
            || || | | || |||
Db   192 GEAPAP----APAEGEAPAPAPAEGEAPAPAP 221

RESULT 13
A27263
50K Spicule matrix protein precursor - sea urchin (Strongylocentrotus purpuratus)
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 19-Nov-1988 #sequence_revision 12-May-1994 #text_change 17-Mar-1999
C;Accession: A27263; A49766
R;Sucov, H.M.; Benson, S.; Robinson, J.J.; Britten, R.J.; Wilt, F.; Davidson, E.H.
Dev. Biol. 120, 507-519, 1987
A;Title: A lineage-specific gene encoding a major matrix protein of the sea urchin embry
A;Reference number: A27263; MUID:87162883
A;Accession: A27263
A;Molecule type: mRNA
A;Residues: 1-199, 'GTSLSANNRAWDXRALVINNQAWVGDNQALVINQEWVGDNQDALVINQEWESDNQAGVINPVVVGDNQAW
                        QVWRDGNQALVINQVWVGDNQAWVASNRIRIRPRTRTRTRITQPGSTDPVCSKRGTRWHNLSSD' <SUC>
A;Note: this sequence has been corrected in reference A49766
R;Katoh-Fukui, Y.; Noce, T.; Deda, T.; Fujiwara, Y.; Hashimoto, N.; Higashinakagawa, T.;
Dev. Biol. 145, 201-202, 1991
A;Title: the corrected structure of the SM50 spicule matrix protein of Strongylocentrotu
A;Reference number: A49766; MUID:91209569
A;Accession: A49766
A;Molecule type: mRNA
A;Residues: 1-445 <KAT>
C;Genetics:
A;Gene: SM50
A;Introns: 35/2
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-445/Product: 50K spicule matrix protein #status predicted <MAT>
F;212-307,321-411/Region: 13-residue repeats (Q-P-G-F/M/W-G-N/G-Q-P-G-V/M-G-G-R)
F;411-430/Region: 3-(or 2-)residue repeats (P-N-N or P-N)

Query Match   21.1%;   Score 214.5; DB 2; Length 445;
Best Local Similarity   34.5%; Pred. No. 1.3e-06;
Matches   77; Conservative   13; Mismatches   54; Indels   79; Gaps   14;

Qy   2 PGFG--VGVGG-IPGVAGVPGVGG-VPGVGGVGVGGVGISPEAQRAAAAAAKXAKYGVGTPAAA 57
           ||| | :|| |:|| || :||:|| ||||||| |:|
Db   227 PGFGNQPGWGGRQPGFGNQPGMGGRQPGWGNQPGVG----------------GRQPGMG------ 269

Qy   58 AAKAAKXAAQFGLVPGVGVAPGV-GVAPGVGVAPGV-GLAPGVGVAPGV------------- 104
           ||||||||||||||| ||| | ||||| | |:|
Db   270 ----------GQQPGWGNQPGVGGRQPGWGGQPGFGNQPGMVDNQAWWTT 318

Qy   105 --------GVAPGVGVAPAIGPPEAQAAAAAKAAKYGVGTPAAAAAAKAAAKAAQPGLVP 155
                    |||||||||| | || :|| |:| |||:| :|
Db   319 RLGNQPGVGGRQPGWGGQPGVG-----------------GRQPGVG------------GRQPGFGNQP 357

Qy   156 GV-GVAPGV-GVAPGVGV-GLAPGV-GVAPGVGVAPGVG 194
           || |||||| ||||| ||| |:|| |||||||||
Db   358 GVGGRQPGWGXGQPGWGGQPGVVGGRQPGMGGRQPGFGNQPGVG 400

RESULT 14
QQBE31
nuclear antigen EBNA1 - human herpesvirus 4
C;Species: human herpesvirus 4, Epstein-Barr virus
C;Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 22-Oct-1999
C;Accession: C43043; S42440; A03773; S33021
R;Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A;Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr vi
A;Reference number: A93065; MUID:85035713
A;Accession: C43043
A;Molecule type: DNA
A;Residues: 1-641 <BAN>
A;Cross-references: EMBL;V01555; NID:g59074; PIDN:CAA24816.1; PID:g1334880
A;Experimental source: strain B95-8
R;Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.;
Nature 310, 207-211, 1984
A;Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A;Reference number: A03794; MUID:84270667
A;Contents: annotation; protein coding region
R;Sample, J.; Hummel, M.; Braun, D.; Birkenbach, M.; Kieff, E.
Proc. Natl. Acad. Sci. U.S.A. 83, 5096-5100, 1986
A;Title: Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear proteins:
A;Reference number: S42440; MUID:86259739
A;Accession: S42440
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-66 <SAM>
A;Cross-references: EMBL;M13941; NID:g330399; PIDN:AAA45889.1; PID:g555157
C;Superfamily: Epstein-Barr virus nuclear antigen
C;Keywords: DNA binding; transcription regulation

Query Match   20.6%;   Score 209.5; DB 1; Length 641;
Best Local Similarity   38.1%; Pred. No. 3.5e-06;
Matches   77; Conservative   1; Mismatches 111; Indels   13; Gaps   7;

Qy   3 GFGVGVGGIPGVAGVPGVGGVPGVGGVGVGGVGISPEAQRAAAAAAKXAKYGVGTPAAAAAKAA 62
           |||||| | ||| |:|| || | | ||| ||||:|||||||
Db   90 GAGAGAGG---AGAGGAGAGGAGAGG--GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 145

Qy   63 AKAAQFGLVPGVGVAPGVGVAPGVTAPGVGL--APGVGVAPGVGVGVAPGVGVAPAIGPP 119
           || || ||||| |||||| | |:|| |||||| ||| ||||||| ||
Db   146 GGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAG 205

Qy   120 EAQAAAAAKAAKYGVGTPAAAAAKAAKAAKAAQPGLVPGVGV-APGVGVAPGVGVAPGVGLA 178
           :||| |||||||:||||||||| ||| |||:| |||||| ||| ||:|||
Db   206 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG---GAGAGGAGGAGGAG-AGGAGGA 261

Qy   179 PGVGVAPGVGVAPGVGVAPAIG 200
           |||||| ||| | ||
Db   262 -GAGGAGGAGGAGGAGGAGG 282

RESULT 15

A70869
hypothetical glycine-rich protein Rv2490c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70869
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: A70869
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1660 <COL>
A;Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA16067.1; PID:g279153
Experimental source: strain H37Rv
netics:
A;Gene: Rv2490c
C;Superfamily: collagen alpha 1(IV) chain

Query Match   20.2%;   Score 205.5; DB 2;   Length 1660; .
Best Local Similarity 35.9%;   Pred. No. 1.3e-05;
Matches   78; Conservative   9; Mismatches 103; Indels 27; Gaps   7;

Qy   3 GPGVGVGGIPGVAGVPGVGGVPGVG---GISPEAQAAAAAKAAKYGVGTPAAA-- 57
         | || | || ||| | ||| |:| || :| | | | | |
Db 222 GAGGGAGGAGGNGNFGHGGAGGVGVGAAGANGANTPGQDGAAGVAGSDDGAGGDGLAGS 281

Qy   58 ------AAKAAAKAAQFGLVPGVVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVG 111
             : | :| |:|| |||| ||| ||||||||||| |: ||||
Db 282 DGGDGAGGVGCNGGRGGWLLGNGGAGGVG---GVGGAGGAGGAGGAGGATGINGPAG 338

Qy 112 VAPAIGPPEAQAAAARKAKAKYGVGTPAAARAKAAKAAQPGLV-----PGVGVAPGVGVA 166
          :: | | ||| | ||| | || || | || || | | |||||| | |
Db 339 ISAAGSDGGAGGNGGA-GGNGGVG---GAGGAGGSAGLLGTVGRAGDGGAGGGGGLGGA 393

Qy 167 PGVGVAPGVG---LAPGVGVAPGVGVAPGVGVAPAIG 200
         ||| | | | || || |:|:|| | | | | |
Db 394 PGDGGAGGNGGGSWLAAGDGGAGGHGGDPGLGGAGGAG 430

rch completed: April 24, 2001, 16:41:54
time: 459 sec

```
            GenCore version 4.5
            Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:          April 24, 2001, 16:42:41 ; Search time 44.88 Seconds
                                            (without alignments)
                                            153.417 Million cell updates/sec

Title:           US-09-340-736-2
Perfect score:   1017
Sequence:        1 FPGPGVGGGIPGVAGVPGV........GVAPGVGVAPGVGVAPGVAPGP 201

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

[ ]ched:         93435 seqs. 3425486 residues

Total number of hits satisfying chosen parameters:   93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	ID	Description
1	696.5	68.5	730	1	ELS_HUMAN	P15502	homo sapien
2	485.5	47.7	747	1	ELS_BOVIN	P04985	bos taurus
3	477	46.9	750	1	ELS_CHICK	P07916	gallus gall
4	412	40.5	864	1	ELS_RAT	Q29372	rattus norv
5	379	37.4	860	1	ELS_MOUSE	P54320	mus musculu
6	256.5	25.2	5263	1	FBOR_BOMMO	P05790	bombyx mori
7	223	21.9	1901	1	Y20B_MYCTU	O53353	mycobacteri
8	216.5	21.3	439	1	XP2_XENLA	P17437	xenopus lae
9	214.5	21.1	445	1	SM5C_STRPU	P11994	strongyloce
10	209.5	20.6	641	1	EBN1_EBV	P03211	epstein-bar
11	206	20.3	1567	1	FMN2_MOUSE	Q91104	mus musculu
12	198.5	19.5	672	1	FHR5_MOUSE	P08399	mus musculu
13	195.5	19.2	407	1	SM41_HEMPU	Q26264	hemicentrot
14	192.5	18.9	627	1	SPD2_NEPCL	P46804	nephila cla
15	188	18.5	743	1	CA18_MOUSE	Q00780	mus musculu
16	186.5	18.3	744	1	CA18_HUMAN	P27658	homo sapien
17	185	18.1	744	1	CA1E_RABIT	P14282	oryctolagus
18	184.5	18.1	1763	1	CA24_ASCSU	P27393	ascaris suu
19	184	18.1	498	1	Y11E_MYCTU	Q50615	mycobacteri
20	182.5	17.9	278	1	YLS6_CAEEL	P34391	caenorhabdi
21	180.5	17.7	157	1	CU63_LOCMI	P45584	locusta mig
22	180.5	17.7	1669	1	CA14_HUMAN	P02462	homo sapien
23	180	17.7	1373	1	CA21_MOUSE	Q01149	mus musculu
24	179.5	17.6	384	1	GRP1_PECHY	P09789	petunia hyb
25	177	17.4	1083	1	T2D3_HUMAN	O00268	homo sapien
26	177	17.4	1685	1	CA54_HUMAN	P29400	homo sapien
27	176.5	17.3	308	1	CC40_CAEEL	P34804	caenorhabdi
28	176	17.3	778	1	YQ34_MYCTU	P71933	mycobacteri
29	175	17.2	145	1	CU65_LOCMI	P45585	locusta mig
30	175	17.2	1356	1	CA21_CANFA	O46392	canis famil
31	174	17.1	1464	1	CA1F_HUMAN	P02452	homo sapien
32	174	17.1	1603	1	CA1F_HUMAN	Q07092	homo sapien
33	173	17.0	152	1	CU64_LOCMI	P11739	locusta mig
34	172.5	17.0	680	1	CA1A_HUMAN	Q03692	homo sapien
35	172	16.9	1364	1	CA21_BOVIN	P03465	bos taurus
36	172	16.9	1460	1	CA1I_CANFA	Q9xSJ7	canis famil
37	170	16.7	1758	1	CA24_CAEEL	P17140	caenorhabdi
38	169	16.6	1669	1	CA14_MOUSE	P02463	mus musculu
39	169	16.6	1747	1	MYSB_ACACA	P19706	acanthamoeb
40	169	16.6	1712	1	CA24_HUMAN	P08572	homo sapien
41	168.5	16.6	747	1	SPD1_NEPCL	P19837	nephila cla
42	168.5	16.6	779	1	CA1I_BOVIN	P02453	bos taurus
43	167	16.4	957	1	Y278_MYCTU	P55877	mycobacteri
44	166.5	16.4	1366	1	CA21_HUMAN	P08123	homo sapien
45	166.5	16.4	1453	1	CA1I_CHICK	P02457	gallus gall

```
                              ALIGNMENTS

RESULT  1
ELS_HUMAN
ID   ELS_HUMAN     STANDARD;      PRT;   730 AA.
AC   P15502;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN).
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin fibroblast;
RX   MEDLINE=89009960; PubMed=3177221;
RA   Fazio M.J., Olsen D.R., Kauh E.A., Baldwin C.T., Indik Z.,
RA   Ornstein-Goldstein N., Yeh H., Rosenbloom J., Uitto J.;
RT   "Cloning of full-length elastin cDNAs from a human skin fibroblast
RT   recombinant cDNA library: further elucidation of alternative splicing
RT   utilizing exon-specific oligonucleotides.";
RL   J. Invest. Dermatol. 91:458-464(1988).
RN   [2]
RP   SEQUENCE OF 603-730 FROM N.A.
RC   TISSUE=Placenta, and Hippocampus;
RX   MEDLINE=96291399; PubMed=8689688;
RA   Frangiskakis J.M., Ewart A.K., Morris C.A., Mervis C.B.,
RA   Bertrand J., Robinson B.F., Klein B.P., Ensing G.J., Everett L.A.,
RA   Green E.D., Proeschel C., Gutowski N.J., Noble M., Atkinson D.L.,
RA   Odelberg S.J., Keating M.T.;
RT   "LIM-kinase1 hemizygosity implicated in impaired visuospatial
RT   constructive cognition.";
RL   Cell 86:59-69(1996).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MOST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   -!- DISEASE: DELETED IN WILLIAMS-BEUREN SYNDROME (WBS), A
CC       DEVELOPMENTAL DISORDER CAUSED BY HAPLOINSUFFICIENCY FOR GENES IN A
CC       2-CM REGION OF CHROMOSOME BAND 7q11.23. HEMIZYGOUS DELETION MAY
CC       CONTRIBUTE TO CERTAIN DEFECTS SUCH AS HYPERCALCEMIA AND GROWTH
CC       DELAY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X15603; CAA33627.1; .
DR   EMBL; M36860; AAA52382.1; .
```

DR   EMBL; U62292; AAB17544.1; -.
DR   PIR; A30524; A30524.
DR   HSSP; P04002; 1WFA.
DR   MIM; 130160; -.
DR   MIM; 194050; -.
KW   Structural protein; Connective tissue; Repeat; Signal.
FT   SIGNAL   1   26
FT   CHAIN   27   730   ELASTIN.
SQ   SEQUENCE   730 AA; 63260 MW; AB06D15BA567AE46 CRC64;

Query Match   68.5%; Score 696.5; DB 1; Length 730;
Best Local Similarity   52.7%; Pred. No. 9.3e-32;
Matches 164; Conservative   6; Mismatches   6; Indels 115; Gaps   7;

Qy   1 FPGFGVGVGGIPGSVAGVPGVGGVPGVGISPEAQAAAAKRAAKYGVGTPAAAAAK   60
          ||||||||||||| || |||||||||||||||||||||||||||||||||||||
Db   404 FPGFGVGVGGIPGSVAGVPGVGGVPGVGISPEAQAAAAKRAAKYGVGTPAAAAAK 463

Qy   61 AAAKAAQF------GLVPGVGTAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAP 114
          |||||||| |||||||||||||||||||||||||||||||||||||||||
Db   464 AAAKAAQFALLNLAGLVPGVGTAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAP 523

Qy   115 AIGPPEAQAA---------   124
          ||| ||
Db   524 GIGPGGVAAAAKSAAKVAAKAQLRAAAGLGAGIPGLGVGVGVGAGYPGLGVGAGV 583

Qy   125 -------AAAKAAKYGVGTP-----------------AAAAKAAAKRAA 149
                 |||||||| ||||||||||||
Db   584 PGPGAVPGALAAAKRAAKYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAKAAAKAARA 643

Qy   150 QFGLVPGVGVAPGVGV--------APGVGVAPGV-------------GLAPGVGVAP--G 186
          |||||| | ||||| || ||| | | |
Db   644 QFGLVGAAGLG-GLGVGGLGVPGVGGLGLGGIPPAAAAKAAKYGAAGLGGVGGAGQPPLGG 702

Qy   187 VGVAPGVGVAP 197
          || |||||||
Db   703 VAARPGPGLSP 713

RESULT 2
ELS_BOVIN
ID   ELS_BOVIN   STANDARD;   PRT;   747 AA.
AC   P04985; P04986; P04987; Q29421;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN).
GN   ELN.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87194772; PubMed=3032943;
RA   Raju K., Anwar R.A.;
RT   "Primary structures of bovine elastin a, b, and c deduced from the
RT   sequences of cDNA clones.";
RL   J. Biol. Chem. 262:5755-5762(1987).
RN   [2]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   TISSUE=Nuchal ligament;
RX   MEDLINE=89274159; PubMed=2543440;
RA   Yeh H., Anderson N., Ornstein-Goldstein N., Bashir M.M.,
RA   Rosenbloom J.C., Abrams W.R., Indik Z., Yoon K., Parks W., Mecham R.,
RA   Rosenbloom J.;
RT   "Structure of the bovine elastin gene and S1 nuclease analysis of
RT   alternative splicing of elastin mRNA in the bovine nuchal ligament.";
RL   Biochemistry 28:2365-2370(1989).
RN   [3]
RP   SEQUENCE OF 1-27 FROM N.A.

RX   MEDLINE=91234332; PubMed=2031719;
RA   Manohar A., Shi W., Anwar R.A.;
RT   "Partial characterization of bovine elastin gene: comparison with the
RT   gene for human elastin.";
RL   Biochem. Cell Biol. 69:185-192(1991).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC   NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC   INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A (SHOWN HERE), B AND C; ARE
CC   PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; J02717; AAA30503.1; -.
DR   EMBL; K03505; AAA30505.1; -.
DR   EMBL; K03506; AAA30506.1; -.
DR   EMBL; J03855; AAA30776.1; -.
DR   EMBL; M58652; AAA03519.1; -.
DR   PIR; A26728; A26728.
DR   PIR; B26728; B26728.
DR   PIR; C26728; C26728.
DR   HSSP; P04002; 1WFA.
KW   Structural protein; Connective tissue; Repeat; Signal;
KW   Alternative splicing.
FT   SIGNAL   1   26
FT   CHAIN   27   747   ELASTIN.
FT   MOD_RES   105   105   OXIDATIVE DEAMINATION.
FT   MOD_RES   109   109   OXIDATIVE DEAMINATION.
FT   MOD_RES   252   252   OXIDATIVE DEAMINATION.
FT   MOD_RES   271   271   OXIDATIVE DEAMINATION.
FT   MOD_RES   275   275   OXIDATIVE DEAMINATION.
FT   MOD_RES   324   324   OXIDATIVE DEAMINATION.
FT   MOD_RES   327   327   OXIDATIVE DEAMINATION.
FT   MOD_RES   400   400   OXIDATIVE DEAMINATION.
FT   MOD_RES   404   404   OXIDATIVE DEAMINATION.
FT   MOD_RES   407   407   OXIDATIVE DEAMINATION.
FT   MOD_RES   448   448   OXIDATIVE DEAMINATION.
FT   MOD_RES   489   489   OXIDATIVE DEAMINATION.
FT   MOD_RES   493   493   OXIDATIVE DEAMINATION.
FT   MOD_RES   544   544   OXIDATIVE DEAMINATION.
FT   MOD_RES   548   548   OXIDATIVE DEAMINATION.
FT   MOD_RES   552   552   OXIDATIVE DEAMINATION.
FT   MOD_RES   606   606   OXIDATIVE DEAMINATION.
FT   MOD_RES   609   609   OXIDATIVE DEAMINATION.
FT   MOD_RES   645   645   OXIDATIVE DEAMINATION.
FT   MOD_RES   649   649   OXIDATIVE DEAMINATION.
FT   MOD_RES   685   685   OXIDATIVE DEAMINATION.
FT   MOD_RES   688   688   OXIDATIVE DEAMINATION.
FT   VARSPLIC   239   239   MISSING (IN ELASTIN B).
FT   VARSPLIC   226   226   MISSING (IN ELASTIN C).
FT   CONFLICT   2   3   RS -> AG (IN REF. 2 AND 3).
FT   CONFLICT   12   12   E -> G (IN REF. 2 AND 3).
SQ   SEQUENCE   747 AA; 64229 MW; 633C033411643D83 CRC64;

Query Match   47.7%; Score 485.5; DB 1; Length 747;
Best Local Similarity   47.8%; Pred. No. 2.5e-20;
Matches 129; Conservative   13; Mismatches   37; Indels   91; Gaps   14;

Qy   5 GVGVGGIPGSVAGVPGVGGVPGVGISPE-AQAAAAAAKAY------GVGTPA   55
          ||||| || |: | | ||||||||||: ||| :|||||||: ||| ||
Db   286 GVGVGG-PGIPGAP--GAIPGIGGIAGVG-APDAAAAAAAAKAAKFGAAGGLPGVGVPG 341

Qy   56 AAAKAAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPA 115

```
Db   342 VGVPGVGVPGVGA---VPGVGV-PGVGV-PGVGV-PGVGV-PGVGV-PGVGVPGA 392
            ||||| |||| ||||| |||| |||| |||: |||||: ||||| |||||
Qy   116 IGP---------------------------PEAQAAAAKAAKY--- 132
            : :                               |: |||||||
Db   393 LSPAATAKAAKAAKSAKFGARGAVGIGGIPTFGLGPGGFPGIGDAAAAPAAAAAKAAKIGAG 452

Qy   133 --------------GVGTPAAAAKAAAKAAQFGLVPGVGVAPGVGVA 166
                          |||||||||||||||| ||||||||||||||
Db   453 GVGAIGGVVPGAPGAIPGLPGVGPGVGVGIPAAAAAKAAANKAAQFGLGPGVGVAPGVGVV 512

Qy   167 PGVVAPGVGLAPGVGVAPGVGVAPGVGVA 196
            |||| ||||:| ||||:|| : |:  |
Db   513 PGVGVPGVGVAPGIGLGPGVGVIGAGVPAA 542
```

RESULT 3

```
CHICK        STANDARD;       PRT;   750 AA.
P07916;
01-AUG-1988 (Rel. 08, Created)
01-AUG-1988 (Rel. 08, Last sequence update)
01-OCT-1996 (Rel. 34, Last annotation update)
ELASTIN PRECURSOR (TROPOELASTIN) (FRAGMENT).
ELN.
Gallus gallus (Chicken).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
Gallus.
NCBI_TaxID=9031;
[1]
SEQUENCE FROM N.A.
MEDLINE=87242320; PubMed=3593675;
Bressan G.M., Argos P., Stanley K.K.;
"Repeating structure of chick tropoelastin revealed by complementary
DNA cloning.";
Biochemistry 26:1497-1503(1987).
[2]
SEQUENCE OF 85-750 FROM N.A.
MEDLINE=88309083; PubMed=2841924;
Baule V.J., Foster J.A.;
"Multiple chick tropoelastin mRNAs.";
Biochem. Biophys. Res. Commun. 154:1054-1060(1988).
[3]
SEQUENCE OF 457-750 FROM N.A.
MEDLINE=87297534; PubMed=3502711;
Tokimitsu I., Tajima S., Nishikawa T., Tajima M., Fukasawa T.;
"Sequence analysis of elastin cDNA from chick aorta and
tissue-specific transcription of the elastin gene in developing chick
embryo.";
Arch. Biochem. Biophys. 256:455-461(1987).
-!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
    NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
-!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
    INTO AN EXTENSIBLE 3D NETWORK.
-!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
-!- ALTERNATIVE PRODUCTS: THERE ARE AT LEAST TWO FORMS OF CHICKEN
    ELASTIN THAT ARE PRODUCED BY ALTERNATIVE SPLICING.
-!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
```

```
EMBL; M18633; AAA48761.1; -.
EMBL; M21880; AAA49082.1; -.
EMBL; M15889; AAA49108.1; -.
PIR; A27264; A27264.
```

```
DR   PIR; A26601; A26601.
DR   BSSP; P04002; IWFA.
KW   Structural protein; Connective tissue; Repeat; Signal;
KW   Alternative splicing;
FT   NON_TER        1      1
FT   SIGNAL        <1     24
FT   CHAIN         25    750                ELASTIN.
FT   DOMAIN        83    686                8 X TANDEM REPEATS.
FT   REPEAT        83    127                1.
FT   REPEAT       219    262                2.
FT   REPEAT       263    318                3.
FT   REPEAT       319    393                4.
FT   REPEAT       394    482                5.
FT   REPEAT       483    554                6.
FT   REPEAT       555    619                7.
FT   REPEAT       620    686                8.
FT   MOD_RES       63     63   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES       66     66   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      111    111   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      115    115   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      156    156   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      159    159   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      198    198   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      200    200   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      235    235   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      252    252   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      256    256   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      297    297   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      301    301   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      354    354   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      357    357   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      427    427   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      431    431   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      513    513   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      517    517   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      520    520   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      586    586   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      590    590   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      593    593   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      655    655   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      658    658   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      719    719   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      721    721   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      743    743   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      748    748   OXIDATIVE DEAMINATION (POTENTIAL).
FT   MOD_RES      750    750   OXIDATIVE DEAMINATION (POTENTIAL).
FT   VARSPLIC     212    212   G -> GLGPGSQOPGVPLGYPIKAPKLPG (IN
FT                             EMBRYONIC ISOFORM).
FT   VARSPLIC     501    501   G -> GVGVPGVGVPG (IN EMBRYONIC ISOFORM).
FT   CONFLICT     536    536   A -> G (IN REF. 3).
FT   CONFLICT     571    571   G -> A (IN REF. 3).
FT   CONFLICT     610    610   A -> A (IN REF. 3).
FT   CONFLICT     654    654   A -> A (IN REF. 3).
FT   CONFLICT     667    667   P -> R (IN REF. 3).
SQ   SEQUENCE   750 AA;  63697 MW;  E57ECD6QC6SE556F CRC64;
```

```
Query Match             46.9%;  Score 477;  DB 1;  Length 750;
Best Local Similarity   58.4%;  Pred. No. 7.2e-20;
Matches 132;  Conservative 12;  Mismatches 46;  Indels 36;  Gaps 19;

Qy     2 PGF---GVGVGGIPGSVAGVPGSVGGVPGVGVGISPEAQAAAAAAKAAKYGVGT---PA 55
           ||    :|:|| :|||| :|||| ||||||| :||:|| ||||| ||||||||    |
Db   363 PGVGVGPGVGIGGVPGGVGVPGVPGVGVGPGVPGVPGVGVPGSVPGVVPGVGGGPAA 422

Qy    56 AAAAKAAAKAAQFGL--VPGVGV--APGVGVAPGVGVAPGVGLAPGVGVAPGVYAPGV 110
           ||||||||:||||:|    ||:||:|  |||||||||:| :|||:|| |||||||||
Db   423 AAAKAAAKAAAFGAGRVPGVGVPGAVPGVGV--PGVGV-PGVGV-PGVGV-PGV 477

Qy   111 GVAPAIGPPEAQAAAAAKAAKYGVGTPA------------AAAKAAAKAAQF---GLVPGV-GV 159
           || || |:  | |||:||:||||||:||            |:||: ||||:     :| : |:
Db   478 GV-PGVGVP---GVGVGVPGSVPGVPGVGVCLVPGAGPAAAAKAAKKAAKYGAGGLAPGVGGL 533
```

```
Qy    160 APGV-GVAPGV-GLAPGV-GVAPGV-GVAPGVGVAPAIG 200
          || |:|:||| |: ||| || || ||||| |:|||||| | :|
Db    534 APAVGGLAPGVGVGGLVPGVGVGGLAPGVGGLAPGVGVGAVPGVG 579


RESULT  4
ID   ELS_RAT          STANDARD;       PRT;    864 AA.
AC   Q99372;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN) (FRAGMENT)
GN   ELN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9104868; PubMed=1702999;
RA   Pierce R.A., Deak S.B., Stolle C.A., Boyd C.D.;
RT   "Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";
RL   Biochemistry 29:9677-9683(1990).
RN   [2]
RP   SEQUENCE OF 781-864 FROM N.A.
RX   MEDLINE=88330866; PubMed=2971041;
RA   Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;
RT   "Elements of the rat tropoelastin gene associated with alternative
RT   splicing.";
RL   J. Biol. Chem. 263:13504-13507(1988).
RN   [3]
RP   SEQUENCE OF 264-533 AND 558-864 FROM N.A.
RX   MEDLINE=92241859; PubMed=1572637;
RA   Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;
RT   "Elements of the rat tropoelastin gene associated with alternative
RT   splicing.";
RL   Genomics 12:651-658(1992).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- ALTERNATIVE PRODUCTS: THREE DIFFERENT MRNAS HAVE BEEN FOUND THAT
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   --------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M60647; AAA42269.1; -.
DR   EMBL; J04035; AAA42268.1; -.
DR   EMBL; M86372; AAA42271.1; -.
DR   EMBL; M86355; AAA42271.1; JOINED.
DR   EMBL; M86363; AAA42271.1; JOINED.
DR   EMBL; M86364; AAA42271.1; JOINED.
DR   EMBL; M86366; AAA42271.1; JOINED.
DR   EMBL; M86371; AAA42271.1; JOINED.
DR   EMBL; M86376; AAA42272.1; -.
DR   EMBL; M86373; AAA42272.1; JOINED.
DR   EMBL; M86375; AAA42272.1; JOINED.
DR   HSSP; P04002; 1WFA.
KW   Structural protein; Connective tissue; Repeat; Signal;
KW   Alternative splicing.
FT   NON_TER.      1      1
FT   SIGNAL       <1     21       BY SIMILARITY.
FT   CHAIN        22    864       ELASTIN.
FT   VARIANT     263    307       MISSING (IN CERTAIN CLONES).
FT   VARIANT     308    308       MISSING (IN CERTAIN CLONES).
FT   VARIANT     809    823       MISSING (IN CERTAIN CLONES).
SQ   SEQUENCE  864 AA;  72786 MW;  456B94BB09E79FD4 CRC64;
```

```
Query Match             40.5%;  Score 412;  DB 1;  Length 864;
Best Local Similarity   47.7%;  Pred. No. 2.6e-16;
Matches 116;  Conservative 12;  Mismatches  43;  Indels  72;  Gaps  18;

Qy     2 PGFGV-GVGGIPGVAGVGVGGTGVGVGGVPGVG----------ISPEAQAAAAAKAAKY- 49
         ||| |:||||| |:||||| ||||||||:||||||:|| :||:|        :||:||||||||
Db   379 PGVGIPGVGGIPGVGGIPGVGG-PGIGGPGIVGGPGAVSPALAAKALAAKAAKYG 437

Qy    50 --GVGTPA-------------------------------AAAKAAAKAAQFGLVPGVGVAPGV-- 80
         || |:|                                  |||| ||||||:||:  | | |:
Db   438 ARGGVGIPYGVGAGGPPGYGVGAGNGLGGASQAAAAAAKAAAKYG-AGGAGTLGGLVP 496

Qy    81 GVAPGV--GVAPGV--GLAPGV--GVAPGVGVAPGVGVAPAIGPPERQAAAAKRAAKYGV 134
         ||||||  || |||  | ||  | ||  |||  |||  ||||       :||||
Db   497 GAVPGGLPGAVPGALPGAVPGALPGAVPGV---PGTGGVPGAGTP----AAAA------ 542

Qy   135 GTPAAAAAAAKAAAQFGLVPGVCVAPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVG 194
         |:||||||||||||| |:|||||| |:||||| ||||  ||  | |:|| |:| |:|:|
Db   543 ---AAAAKAAAKAGQYGLGPGVGPVG-GV--GVGGLPG-GVGPG-GVT-GIGTGPGTG 593

Qy   195 VAP 197
         |:|
Db   594 LVP 596


RESULT  5
ID   ELS_MOUSE        STANDARD;       PRT;    860 AA.
AC   P54320;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN).
GN   ELN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Lung;
RX   MEDLINE=95130069; PubMed=7829060;
RA   Wydner K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT   "Use of an intron polymorphism to localize the tropoelastin gene to
RT   mouse chromosome 5 in a region of linkage conservation with human
RT   chromosome 7.";
RL   Genomics 23:125-131(1994).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   --------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U08210; AAA80155.1; -.
DR   HSSP; P04002; 1WFA.
DR   MGD; MGI:95317; Eln.
KW   Structural protein; Repeat; Signal; Connective tissue.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    860       ELASTIN.
SQ   SEQUENCE  860 AA;  71955 MW;  0COBE5AAE1EDD7F1 CRC64;
```

Query Match   37.3%; Score 379; DB 1; Length 860;
Best Local Similarity   38.3%; Pred. No. 1.6e-14;
Matches 128; Conservative 128; Mismatches   57; Indels 138; Gaps   20;

Qy   2 PGFGVGVGIPGVA------GVP------------------------------GVGGVPGVGGV 29
              |||| |||| ||| || ||||||
Db   243 PGGGGAPSGIPGVGPFGSGQQPGVPLGYPLPIRAPKLPGGYGLPYTNGKLPYGVRAGGRAGY 302

Qy   30 P-GVGISPEAQAAAAAKAAKY---------------------------------GVGTP-AAA 57
             | |::: ||||||||||| ||||| |||
Db   303 PTGTGVGSQA-AAAAAKAAKYGAGGAGVLPGVGGGIPGSAGAIPGIGGIGAGGTPAAAA 361

Qy   58 AAKAAKAAQF-----GIVP-GVGV-APGVGVAPGVGVAPGVGLAPSVGVAPGVGVAPGVG 111
             ||||||||| |||| |||| || | |||| |||: ||: |||: ||
Db   362 AAKAAKARAKYGAAGGIVPGGPGVRLPGAGI-PSVGGIPGVGG-GPGIG-GPGIV 418

Qy   112 VAPAIGPPEAQAAAAAKAAKY----GVGTP------------------------AAAAKA 144
             |||:|||||| ||||||||||| |||| |||||
Db   419 GGPGAVSPAAAAAKAAAKAAKYGARGGVGIPFYGVGAGGFPGYGVGAGAGLGGAASPAAAAA 478

Qy   145 AAKAAQP---------GIVPGV-GYAPG-VGVAPGVGVAPG--------------------- 174
             |:||| ||| | | | ||
Db   479 AAKAAKYGAGGAGALGGLVPGAVPAVPGAGGVPGAGVPGAAAAAAAAKAAAK 538

Qy   175 VGLAPGVGVAP----GVGVAPG----VGVAPAIGP 201
             || ||||||| ||| || ||||||||
Db   539 AGLGPGVGVEGGVGVGGIPGGVGVGGVGVPGGVGP 572

RESULT 6
FBOH_BOMMO
ID   FBOH_BOMMO   STANDARD;   PRT; 5263 AA.
AC   P05790; Q26379; Q17220;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   FIBROIN HEAVY CHAIN PRECURSOR (FIB-H) (H-FIBROIN).
GN   FIBH.
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20303362; PubMed=10871375;
RA   Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Esnault C.,
RA   Yang T., Jacquet M., Janin J., Duguet M., Perasso R., Li Z.-G.;
RT   "Fine organization of Bombyx mori fibroin heavy chain gene.";
RL   Nucleic Acids Res. 28:2413-2419(2000).
RN   [2]
RP   SEQUENCE OF 1-168 FROM N.A.
RX   MEDLINE=80045039; PubMed=498286;
RA   Tsujimoto Y., Suzuki Y.;
RT   "The DNA sequence of Bombyx mori fibroin gene including the 5'
RT   flanking, mRNA coding, entire intervening and fibroin protein coding
RT   regions.";
RL   Cell 18:591-600(1979).
RN   [3]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=79211211; PubMed=455439;
RA   Tsujimoto Y., Suzuki Y.;
RT   "Structural analysis of the fibroin gene at the 5' end and its
RT   surrounding regions.";
RL   Cell 16:425-436(1979).
RN   [4]
RP   PARTIAL SEQUENCE FROM N.A.
RC   STRAIN=KINSHU X SHOWA;
RX   MEDLINE=89046668; PubMed=3210244;
RA   Mita K., Ichimura S., Zama M., James T.C.;
RT   "Specific codon usage pattern and its implications on the secondary

RT   structure of silk fibroin mRNA.";
RL   J. Mol. Biol. 203:917-925(1988).
RN   [5]
RP   PARTIAL SEQUENCE FROM N.A.
RX   MEDLINE=94365842; PubMed=7916056;
RA   Mita K., Ichimura S., James T.C.;
RT   "Highly repetitive structure and its organization of the silk fibroin
RT   gene.";
RL   J. Mol. Evol. 38:583-592(1994).
RN   [6]
RP   SEQUENCE OF 5179-5263 FROM N.A. AND DISULFIDE BONDS.
RC   STRAIN=J-139;
RX   MEDLINE=93296390; PubMed=103667;
RA   Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,
RA   Takegi T., Mizuno S.;
RT   "Determination of the site of disulfide linkage between heavy and
RT   light chains of silk fibroin produced by Bombyx mori.";
RL   Biochim. Biophys. Acta 1432:92-103(1999).
CC   -!- FUNCTION: FORMS THE SILK FILAMENT: A STRONG, INEXTENSIBLE,
CC   INSOLUBLE AND CHEMICALLY INERT FIBRE.
CC   -!- SUBUNIT: FORMED OF TWO CHAINS: HEAVY AND LIGHT, THAT ARE LINKED BY
CC   A DISULFIDE BOND. HEAVY-LIGHT CHAIN ASSEMBLY IS ESSENTIAL FOR THE
CC   EFFICIENT INTRACELLULAR TRANSPORT AND SECRETION OF FIBROIN.
CC   -!- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE POSTERIOR (PSG)
CC   SECTION OF SILK GLANDS.
CC   -!- DOMAIN: COMPOSED OF ANTIPARALLEL BETA SHEETS. THE STRANDS OF THE
CC   BETA SHEETS RUN PARALLEL TO THE FIBER AXIS. LONG STRETCHES OF SILK
CC   FIBROIN ARE COMPOSED OF MICROCRYSTALLINE ARRAYS OF (-GLY-SER-GLY-
CC   ALA-GLY-ALA-)N INTERRUPTED BY REGIONS CONTAINING BULKIER RESIDUES.
CC   THE FIBER IS COMPOSED OF MICROCRYSTALLINE ARRAYS ALTERNATING WITH
CC   AMORPHOUS REGIONS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AF226688; AAF76983.1; -.
DR   EMBL; V00094; CAA23432.1; -.
DR   EMBL; V00097; CAA23433.1; -.
DR   EMBL; S74439; AAB31861.1; -.
DR   EMBL; X13069; CAA32076.1; -.
DR   EMBL; M35378; AAA27839.1; -.
DR   EMBL; AB017362; BAA33147.1; -.
DR   PIR; S01844; S01844.
KW   Silk; Signal; Repeat.
FT   SIGNAL   1   21
FT   CHAIN   22   5263   FIBROIN HEAVY CHAIN.
FT   DOMAIN   149   5206   HIGHLY REPETITIVE.
FT   DISULFID 5244   5244   INTERCHAIN (WITH LIGHT CHAIN).
FT   DISULFID 5260   5263
FT   CONFLICT   10   10   C -> V (IN REF. 2).
SQ   SEQUENCE   5263 AA; 391586 MW; 8EE1ID3A0A474408 CRC64;

POTENTIAL

Query Match   25.2%; Score 256.5; DB 1; Length 5263;
Best Local Similarity   32.9%; Pred. NO. 2.5e-07;
Matches   69; Conservative   5; Mismatches 121; Indels   15; Gaps   3;

Qy   6 VGVGGIPGVAGV-2GVGVPGVGGVPGVPGVGISPEAQA------AAAAKAAKYGVGTPAAA- 57
              || || | | || |||| ||| ||||||||| ||||||||||||||
Db   151 VGAGAGAAASGAGAGAGTGAASGAGAGAGAGAGAGAGAGTGAGAGAGAGAGAGAGAGAG 210

Qy   58 ----AAKAAKAAQFGLVPGVVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVA 113
                 |:: ::: |||||| |||||| |||||||||| || |||||
Db   211 AGYGAGAGAGCAGYGAGAGAGCAGYGAGAGAGAGAGAGCYGAGAGAGCYGAAGAGAG 270

Qy   114 PAIGPPEAQAAAAAKAAKYGVGTPAAAAAKAAA---KAAQFGLVPGVGVAPGVGPGVG 170
              |: |||| ||||||| |||| | |||| ||: | | ||| ||| |
Db   271 AGYGCGVGSCAASGAGAGAGSAAGSGAGCAGYGAGAGCYGAGAGAGCYGAASGVG 330

Qy   171   VAPGVGLAPGVGVAPGVGVAPAIG 200
            | : | | | | | | | | | | | :|:
Db   331   AGYGAGAGAGTGGASGAGGAGGAGAGAGAG 360

```
RESULT 7
ID   YZ08_MYCTU    STANDARD;      PRT;  1901 AA.
AC   O53553;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.
GN   RV3508 OR MTV023.15.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton J., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL022022; CAA17745.1; -.
DR   HSSP; P19972; 1KVD.
DR   TubercuList; Rv3508; -.
DR   InterPro; IPR000084; -.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein; Repeat; Signal.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1901       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                               RV3508.
SQ   SEQUENCE   1901 AA;  147627 MW;  C7B1923D5D0146CD CRC64;
```

Query Match   21.9%;   Score 223;   DB 1;   Length 1901;
Best Local Similarity   32.9%;   Pred. No. 7.9e-06;
Matches   77;   Conservative   7;   Mismatches 114;   Indels   36;   Gaps   4;

Qy   3   GPGV--GVGGIPGVAGTPGVGPGVGGVPGVGISPEAQAAAAKAAKYGVGTPAAAAAK 60
         | || ||| | |||| :| || |||| | ||| | |
Db   340   GDSVLGSVGGSTGGKGGVGSVVAGLGGAGGAGAGQLFSAGGAGGAGAGVVGGCTGGQGGAGGAGAA 399

Qy   61   AAAKAAQPGLVPGVGVAPGVG-------VAPGVGVAPGVGLAPGVGVAPGVG------ 105
         ||| ||||||| |:| | | |||| |:|:
Db   400   GADAPASTGLRGGTGTGGTGFAGGAGGVGGQGGNAIAGGINGSGGAGGTGGQGGAGGMGGSGADN 459

Qy   106   ------------------VAPGVGVAPAIGPPEAQAAAKAAAAKYGVGTP--AAAAAKAAA 146
         ||| || |:| ||| | ||||
Db   460   ASSIGADGAGGAGGTGGNACAGAGGGAGGAAGTGGTGGVVGAAGKAGIGGTGGQGGAGGSAGT 519

Qy   147   KAAQPGLVPGVGVAPGVGVAPGVGVAPGVGVGLAPGVGVAPGVGVAPGVGVAPAIG 200
         | | | || | | | | | | | | | | | :| |:| | | | | :|
Db   520   DNIATGATGGTGFSGGAGGAGGAGGNTSVGGTNGSGGQGGTGCAGGAGGVG 573

```
RESULT 8
ID   XP2_XENLA     STANDARD;      PRT;   439 AA.
AC   P17437; Q08944;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   SKIN SECRETORY PROTEIN XP2 PRECURSOR (APEG PROTEIN).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE OF 1-25 AND 344-439 FROM N.A.
RX   MEDLINE=92333564; PubMed=1629230;
RA   Hauser F., Roeben C., Hoffmann W.;
RT   "XP2, a new member of the P-domain peptide family of potential growth
RT   factors, is synthesized in Xenopus laevis skin.";
RL   J. Biol. Chem. 267:14451-14455(1992).
RN   [2]
RP   SEQUENCE OF 3-439 FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=90127399; PubMed=2298293;
RA   Gmachl M., Berger H., Thalhammer J., Kreil G.;
RT   "Dermal glands of Xenopus laevis contain a polypeptide with a highly
RT   repetitive amino acid sequence.";
RL   FEBS Lett. 260:145-148(1990).
CC   -!- FUNCTION: MAY ACT AS A GROWTH FACTOR IN THE GERMINAL LAYER OF THE
CC       EPIDERMIS. MAY ALSO BE INVOLVED IN GROWTH OF REGENERATING GLANDS
CC       AND IN PROTECTION OF THE SKIN FROM THE EXTERNAL ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS OF THE PROTEINS ARE
CC       PRODUCED BY ALTERNATIVE SPLICING. THE APEG FORM INCLUDES A LARGE
CC       DOMAIN WITH 33 SEQUENCE REPEATS.
CC   -!- TISSUE SPECIFICITY: SKIN.
CC   -!- SIMILARITY: CONTAINS 2 P-TYPE (TREFOIL) DOMAINS.
CC   -!- CAUTION: REP.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 392
CC       ONWARD AND IS SHORTER (418 AA) DUE TO A FRAMESHIFT.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M90095; AAA50001.1; -.
DR   EMBL; X51394; CAA35759.1; ALT_FRAME.
DR   PIR; S07498; SKKLAG.
DR   PIR; A37331; A37331.
DR   HSSP; P04155; 1PS2.
DR   InterPro; IPR000519; -.
DR   Pfam; PF00088; trefoil; 2.
DR   PRINTS; PR00680; PTREFOIL.
DR   PROSITE; PS00025; P_TREFOIL; 2.
KW   Signal; Growth factor; Alternative splicing; Amphibian skin; Repeat.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    439       SKIN SECRETORY PROTEIN XP2.
FT   MOD_RES      23     23       PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT   DOMAIN       26    343       33 X REPEATS OF G-[GE]-[AP](2,4)-A-E.
FT   DOMAIN      350    391       P-TYPE 1.
FT   DOMAIN      397    438       P-TYPE 2.
FT   DISULFID    351    377       BY SIMILARITY.
FT   DISULFID    361    376       BY SIMILARITY.
FT   DISULFID    371    388       BY SIMILARITY.
FT   DISULFID    398    424       BY SIMILARITY.
```

```
FT   DISULFID    408    423    BY SIMILARITY.
FT   DISULFID    418    435    BY SIMILARITY.
FT   VARSPLIC    26     343    MISSING (IN ISOFORM XP2).
FT   CONFLICT    3      3      H -> S (IN REF. 2).
FT   CONFLICT    18     18     C -> W (IN REF. 2).
SQ   SEQUENCE    439 AA;  41173 MW;  38C4A4B57C8AE778 CRC64;
```

```
Query Match             21.3%;  Score 216.5;  DB 1;  Length 439;
Best Local Similarity   36.0%;  Pred. No. 6.3e-06;
Matches  77;  Conservative   6;  Mismatches 102;  Indels  29;  Gaps    7;

Qy    5 GVGVGIP---GVAGVPGVGGVP------GVGGVPGVGISPEAQARAAAAAKAAKYGVGTP 54
          | |     ||  | |   ||         |  |  ||||  :|:   |  |||    |
Db   22 GQDAGGEAPAEGVAPAPAEGGAPAPAPAEGERAPAPAEGGAPAPAEGAPAPAEGAPAPDGGAP 81

Qy   55 AAAAKAAKAAAQFGLVPGVVAPGVGAPGVGVAPGVGLAPGVGAPG--VGVAPGVGV 112
          ||||       :|  |     | ||    | | ||||  | | |||   ||    ||
Db   82 APAPAEGGAPAP---APAEGGAPAPAEGGAPAPAEGGAPAPAEGGAPAPAEGEAPAPAEGE 137

Qy  113 APAIGPPEAQAAAAAKA----AKYGVGTPAAAAAKAAKAAAQFGLVPGVGVAPGVGVAP 167
          || |           |     ||: |                  |:||  ||  |
Db  138 APAPAEGGAPAPAEGGAPAPAEGGAPAPAEGAPAPAEGGAPAP----APAEGEAPAPAPAP 193

Qy  168 GVGVAPGVGLAPGVGGVAPGVGVAPGVGVAPGVGVAPAIGP 201
          |  |  |   ||  || | || | | |
Db  194 GEAPAP----APAEGEAPAPAPAPAEGGEAPAPAP 223
```

```
RESULT  9
SM50_STRPU
ID   SM50_STRPU   STANDARD;     PRT;    445 AA.
AC   P11994;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   50 KDA SPICULE MATRIX PROTEIN PRECURSOR.
GN   SM50.
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
OX   NCBI_TaxID=7668;
RN   [1]
RP   SEQUENCE FROM N.A.
```
```
RX   MEDLINE=87162883; PubMed=3030858;
RA   Sucov H.M., Benson S.C., Robinson J.J., Britten R.J., Wilt F.H.,
RA   Davidson E.H.;
RT   "A lineage-specific gene encoding a major matrix protein of the sea
RT   urchin embryo spicule. II. Structure of the gene and derived sequence
RT   of the protein.";
RL   Dev. Biol. 120:507-519(1987).
```
```
RN   [2]
RP   REVISIONS.
```
```
RX   MEDLINE=91209669; PubMed=2019323;
RA   Katoh-Fukui Y., Noce T., Ueda T., Fujiwara Y., Hashimoto N.,
RA   Higashinakagawa T., Noll M., Livingston B.T., Wilt F.H.,
RA   Benson S.C., Sucov H.M., Davidson E.H.;
RT   "The corrected structure of the SM50 spicule matrix protein of
RT   Strongylocentrotus purpuratus.";
RL   Dev. Biol. 145:201-202(1991).
```
```
CC   -!- FUNCTION: MAJOR MATRIX PROTEIN OF THE SEA URCHIN EMBRYO SPICULE.
CC       THE FUNCTION OF THE MATRIX PROTEINS IS TO DIRECT CRYSTAL GROWTH IN
CC       CERTAIN ORIENTATIONS AND INHIBIT GROWTH IN OTHERS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE
CC       MICROMERE/PRIMARY MESENCHYME CELLS (PMC) LINEAGE.
CC   -!- DEVELOPMENTAL STAGE: DETECTED AT HATCHING BLASTULA. IT REMAINS AT
CC       A CERTAIN LEVEL FROM THE MESENCHYME BLASTULA TO THE PLUTEUS
CC       STAGES.
CC   -!- DOMAIN: THE REPETITIVE DOMAIN MAY PROVIDE A CALCITE BINDING
CC       MATRIX.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE SM50 FAMILY.
```

```
DR   EMBL; M16231; AAA30071.1; -.
DR   EMBL; M16230; AAA30071.1; JOINED.
DR   PIR; A27263; A27263.
DR   InterPro; IPR001304; -.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
DR   PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
KW   Matrix protein; Repeat; Signal.
FT   SIGNAL        1     15
FT   CHAIN        16    445        50 KDA SPICULE MATRIX PROTEIN.
FT   DOMAIN        ?     ?         C-TYPE LECTIN.
FT   DOMAIN      123    180        PRO-RICH.
SQ   SEQUENCE    445 AA;  46262 MW;  1DD08FD0882639A CRC64;
```

```
Query Match             21.1%;  Score 214.5;  DB 1;  Length 445;
Best Local Similarity   34.5%;  Pred. No. 8.1e-06;
Matches  77;  Conservative  13;  Mismatches 79;  Indels  14; Gaps 14;

Qy    2 PGFG--VGVGG-IPGVAGVPGVGG-VPGVGGVPGVGISPEAQAAAAAKAAKYGVGTPAAA 57
          ||||       |||  || ||    || |  ||||        :|:|
Db  227 PGFGNQPGMGGRQPGFGNQPGMGGRQPGWGNQPGVG---------GRQPGMG------- 269

Qy   58 AAKAAKAAAQFGLVPGVGVAPGV-GVAPGVGVAPGV-GLAPGVGVAPGV----------- 104
          ||  |       ||  |  ||  |  ||   ||  |    | |  ||
Db  270 -------GQQPGNQPGVGGRQPGMGGQPGVGGRQPGFGNQPGMVDNNQAWWTTT 318

Qy  105 -----------GVAPGVGVAPAIGPPEAQAAAAAKAAKYGVGTPAAAAAKAAAQFGLVP 155
                       ||       |||        :|||
Db  319 RLGNQPGVGGRQPGMGGQPGVG----------GRQPGVG--------GRQPGFGNQP 357

Qy  156 GVGVAPGV-GVAPGVGVAPGV-GLAPGV-GVAPGVGVAPGVG 194
          | || |  |   ||  |||  |  |  ||| | || |
Db  358 GVGGRQPGMGGQPGVGGQPGVGGRQPGMGGRQPGFGNQPGVG 400
```

```
RESULT  10
EBN1_EBV
ID   EBN1_EBV     STANDARD;     PRT;    641 AA.
AC   P03211;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   EBNA-1 NUCLEAR PROTEIN.
GN   BKRF1.
OS   Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Lymphocryptovirus.
OX   NCBI_TaxID=10377;
RN   [1]
RP   SEQUENCE FROM N.A.
```
```
RX   MEDLINE=84270667; PubMed=6087149;
RA   Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA   Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA   Tuffnell P.S., Barrell B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   Nature 310:207-211(1984).
```
```
RN   [2]
RP   SEQUENCE OF 1-26 FROM N.A.
```
```
RX   MEDLINE=86259739; PubMed=3460083;
RA   Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;
RT   "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear
RT   proteins: a probable transcriptional initiation site.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).
```
```
RN   [3]
```

RP   SUBCELLULAR LOCATION.
RX   MEDLINE=90266473; PubMed=2161150;
RA   Petti L., Sample C., Kieff E.;
RT   "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT   latent infection nuclear proteins.";
RL   Virology 176:563-574(1990).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.
RX   MEDLINE=96006523; PubMed=7553871;
RA   Bochkarev A., Barwell J.A., Pfuetzner R.A., Furey W.F. Jr.,
RA   Edwards A.M., Frappier L.;
RT   "Crystal structure of the DNA-binding domain of the Epstein-Barr
RT   virus origin-binding protein EBNA 1.";
RL   Cell 83:39-46(1995).
CC   -!- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE
CC   MAINTENANCE REPLICATION OF EBV EPISOME. TRANSACTIVATING FACTOR
CC   FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORIP.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOPLASM, SOMEWHAT
CC   ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED
CC   WITH THE NUCLEAR MATRIX.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; V01555; CAA24816.1; -.
DR   EMBL; M13941; AAA45889.1; -.
DR   PIR; A03773; QQBE31.
DR   PIR; S33021; S33021.
DR   PDB; 1VHI; 23-DEC-95.
DR   TRANSFAC; T00211; -.
KW   Nuclear protein; DNA-binding; Transcription regulation; Activator;
KW   3D-structure.
FT   DOMAIN   87   352   GLY/ALA-RICH.
SQ   SEQUENCE   641 AA;   56427 MW; 4D16I653EI6PC341 CRC64;

Query Match   20.6%; Score 209.5; DB 1; Length 64;
Best Local Similarity   38.1%; Pred. No. 2e-05;
Matches   77; Conservative   1; Mismatches 111; Indels 13; Gaps 7;

Qy   3 GFGVGVGGIPGVANGVPGVGGVPGGVGVPGVGISPEAQAaaaaAkAKYGVGTPAaaAAKAA 62
Db   90 GAGAGAGG--AGAGGAGGGGAGGAGG--GAGGAGGGAGGAGGGAGGAGGAGGAGAGAG 145

Qy   63 AKAAAQPGLVPGVGVAPGVGVAPGVGL--APGVGVAPGVGVAPGVGVAPGVGPP 119
Db 146 GGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGGAGGGAGAG 205

Qy 120 EAQAAAAKAAKYGVGTPAAAAAKAAKAAAQFGLVPGVGV-APGVGVAPGVGVAPGVGLA 178
Db 206 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG---GAGGAGGAGGAGGAG-AGGAGGA 261

Qy 179 PGVGVAPGVGVAPGVGVAPAIG 200
Db 262 -GAGGAGGAGGAGGGAGGAGAGG 282

RESULT 11
FMN2_MOUSE
ID   FMN2_MOUSE   STANDARD;   PRI;   1567 AA.
AC   Q9JL04;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   FORMIN 2.
GN   FMN2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20245324; PubMed=10781961;
RA   Leader B., Leder P.;
RT   "Formin-2, a novel formin homology protein of the cappuccino
RT   subfamily, is highly expressed in the developing and adult central
RT   nervous system.";
RL   Mech. Dev. 93:221-231(2000).
CC   -!- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN THE DEVELOPING
CC   AND MATURE CENTRAL NERVOUS SYSTEM.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSION BEGINS AT EMBRYONIC DAY 9.5 IN
CC   THE DEVELOPING SPINAL CORD AND BRAIN STRUCTURES AND CONTINUES IN
CC   NEONATAL AND ADULT BRAIN STRUCTURES INCLUDING THE OLFACTORY BULB,
CC   CORTEX, THALAMUS, HYPOTHALAMUS, HIPPOCAMPUS AND CEREBELLUM.
CC   -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 1 (FH1) DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 2 (FH2) DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE FORMIN HOMOLOGY FAMILY. CAPPUCCINO
CC   SUBFAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AF218940; AAF72883.1; -.
DR   MGD; MGI:1855252; Fmn2.
DR   InterPro; IPR0001265; -.
DR   InterPro; IPR003104; -.
DR   Pfam; PF02181; FH2; 1.
DR   PRINTS; PR00628; FORMIN.
KW   Developmental protein; Repeat; Coiled coil.
FT   DOMAIN   643   683   COILED COIL (POTENTIAL).
FT   DOMAIN   735   1113   FH1 (PRO-RICH).
FT   DOMAIN   919   1039   11 X 11 AA TANDEM REPEATS OF [MV]-G-I-P-
FT   P-P-P-L-P-G.
FT   REPEAT   919   929   1.
FT   REPEAT   930   940   2.
FT   REPEAT   941   951   3.
FT   REPEAT   952   962   4.
FT   REPEAT   963   973   5.
FT   REPEAT   974   984   6.
FT   REPEAT   985   995   7.
FT   REPEAT   996   1006   8.
FT   REPEAT   1007   1017   9.
FT   REPEAT   1018   1028   10.
FT   REPEAT   1029   1039   11.
FT   DOMAIN   1128   1532   FH2.
FT   DOMAIN   1408   1444   COILED COIL (POTENTIAL).
FT   DOMAIN   48   55   POLY-GLY.
FT   DOMAIN   202   207   POLY-GLN.
FT   DOMAIN   797   801   POLY-PRO.
FT   DOMAIN   861   864   POLY-PRO.
FT   DOMAIN   908   917   POLY-PRO.
FT   DOMAIN   922   928   POLY-PRO.
FT   DOMAIN   933   939   POLY-PRO.
FT   DOMAIN   944   950   POLY-PRO.
FT   DOMAIN   955   961   POLY-PRO.
FT   DOMAIN   966   972   POLY-PRO.
FT   DOMAIN   977   983   POLY-PRO.
FT   DOMAIN   988   994   POLY-PRO.
FT   DOMAIN   999   1005   POLY-PRO.
FT   DOMAIN   1010   1016   POLY-PRO.
FT   DOMAIN   1021   1027   POLY-PRO.
FT   DOMAIN   1032   1038   POLY-PRO.
FT   DOMAIN   1043   1046   POLY-PRO.
FT   DOMAIN   1054   1060   POLY-PRO.
FT   DOMAIN   1065   1072   POLY-PRO.

```
FT   DOMAIN    1077   1080   POLY-PRO.
SQ   SEQUENCE  1557 AA; 162268 MW; 8F2173B1C8E505944 CRC64;

Query Match       20.3%;  Score 206; DB 1; Length 1557;
Best Local Similarity  30.9%;  Pred. No. 5.7e-05;
Matches   69;  Conservative  21;  Mismatches  55;  Indels  78;  Gaps  14;

Qy     5  GVGVGGIPGVAGVPGVGGVPGVGGVPGVGISPEAQAAAAAKAAAKYGVGTPAAAAAKAAAK  64
                     |::|| |     :||:|| :||:| |  |:|
Db   904  GLGVPEPPAPPLPGM-GIPPPPLPGMGIPPPPLP--------GMGIPPPP----------  948

Qy    55  AAQFGLVPGVGV----APGVGV-----APGVGVGLAPGVGV----APGVGV---  106
              ::||||       ||||::        ||||    |||||
Db   949  --LPGVGIPPPPLPGVGIPPPPLPGVGIPPPPLPGVGIPPPPLPGVGIPPPP  1001

Qy   107  --APGVGVAPAIGPPEAQAAAAAKAAAKYGVGTPAAAAAKAAAKAAQTGLVPGVGV---  159
              ||||                             ||||
Db  1002  PPLPGVGIPP----PPPLP--------GVGIPPPP----------LPGXGIPPPP  1036

Qy   160  APGVGV----APGVGVGLAPGVGVGLAPGVGVAPGVGVGVAPGVGVAPGVGVGVAPGVGVAP  197
              ||||       |:|:|||
Db  1037  LPGSSGIPPPPALPGVAIPPPPPL-PGMGVPPPAPPPPGAGIPP  1078
```

RESULT 12

```
ID   PHX5_MOUSE      STANDARD;       PRT;    672 AA.
AC   P03399;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PHX5 OR PER.
DE   PER-HEXAMER REPEAT PROTEIN 5.
GN   PHX5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86014384; PubMed=2413365;
RA   Shin H.S., Bargiello T.A., Clark B.T., Jackson F.R., Young M.W.;
RT   "An unusual coding sequence from a Drosophila clock gene is conserved
RT   in vertebrates.";
RL   Nature 317:445-448(1985).
CC   -!- PPM: THE SERINE RESIDUES OF THE S-G REPEATS FOUND IN CERTAIN
CC       PROTEOGLYCANS ARE ATTACHMENT SITES FOR GLYCOSAMINOGLYCANS.
CC       ALTHOUGH IT IS NOT KNOWN WHETHER THIS PROTEIN BINDS PROTEOGLYCAN.
CC   -!- SIMILARITY: TO OTHER G-T STRETCH CONTAINING PROTEINS.
CC
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
DR   EMBL; M12039; AAA89320.1; -.
DR   EMBL; X02966; CAA26710.1; ALT_INIT.
DR   PIR; A24403; UMMS.
DR   MGD; MGI:104521; Phx5.
DR   InterPro; IPR000561; -.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW   Proteoglycan; Repeat.
FT   DOMAIN     59    672    G-T REPEATS.
SQ   SEQUENCE  672 AA; 57924 MW; B85BF428CF424C0B CRC64;

Query Match       19.5%;  Score 198.5; DB 1; Length 672;
Best Local Similarity  28.2%;  Pred. No. 8e-05;
Matches   59;  Conservative   3;  Mismatches 128;  Indels 128;  Gaps   4;
```

```
Qy     3  GRGVGVGGIPGVAGVPGVGGVPGVGGVPGVGI------SPEAQAAAAKAAKYGVGTPA  55
              |||||||  :| ::|: || :||:|| |           |:  || ||| ||||||
Db   152  GTGTGTGTGTGTAKVTGTGDRGTGTGTGTGTGTGTGTCGTGTGTAKVTGTAKVTGTGTGT--  209

Qy    56  AAAAKAAAKAAQFGLVPGVGVGVAPGVGVAPGVGVGVA---PGVGLAPGVGVAPGVGVAPGVG  111
              |:|  |||:||                                ||  |||| |||| || |
Db   210  ----AKVTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTAKVTGTGTGTGTGTGTGTGTGTGTG  264

Qy   112  VAPAIGPEAQAAAAAKAAKYGVSTPAAAAAKAAAKAAQFLVPGVGVAPGVGV  171
              || |  || | || | |:|  || |||  |||| ||| |:
Db   265  TAKVTGTGIDRGIGTGTGTGTGTGT-GTGTGTGTAKVTGTGTGTGTGTGTGTGTGTGTGT  323

Qy   172  APGVGLAPGVGAPGVGVAPGVGVAPAIG  200
              ||||  |:|| ||| ||| ||||:|
Db   324  GTGTGTGSGSGTGTGTGTGSGSGTGTGTG  352
```

RESULT 13

```
ID   SM41_HEMPU      STANDARD;       PRT;    407 AA.
AC   Q26234;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   41 KDA SPICULE MATRIX PROTEIN PRECURSOR (HSM41) (HPSMC).
OS   Hemicentrotus pulcherrimus (Sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Hemicentrotus.
OX   NCBI_TaxID=7650;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93075615; PubMed=1445780;
RA   Katoh-Fukui Y., Noce T., Ueda T., Fujiwara Y., Hashimoto N.,
RA   Tanaka S., Higashinakagawa T.;
RT   "Isolation and characterization of cDNA encoding a spicule matrix
RT   protein in Hemicentrotus pulcherrimus micromeres.";
RL   Int. J. Dev. Biol. 36:353-361(1992).
CC   -!- FUNCTION: MAJOR MATRIX PROTEIN OF THE SEA URCHIN EMBRYO SPICULE
CC       WHICH DIRECTS CRYSTAL GROWTH IN CERTAIN ORIENTATIONS AND INHIBIT
CC       GROWTH IN OTHERS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE
CC       MICROMERE/PRIMARY MESENCHYME CELLS (PMC) LINEAGE.
CC   -!- DEVELOPMENTAL STAGE: APPEARS IN GASTRULAE AND REMAINS AT A SIMILAR
CC       LEVEL UNTIL THE PLUTEUS STAGE.
CC   -!- DOMAIN: THE REPETITIVE DOMAIN MAY PROVIDE A CALCITE BINDING
CC       MATRIX.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE SM50 FAMILY.
CC
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
DR   EMBL; S48755; AAB24285.1; -.
DR   InterPro; IPR001304; -.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
DR   PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
KW   Matrix protein; Signal; Repeat.
FT   SIGNAL      1     17       POTENTIAL.
FT   CHAIN      18    407       41 KDA SPICULE MATRIX PROTEIN.
FT   DOMAIN    104      ?       C-TYPE LECTIN.
FT   DOMAIN    179              PRO-RICH.
SQ   SEQUENCE  407 AA; 42383 MW; 5A5397C46B80F685 CRC64;

Query Match       19.2%;  Score 195.5; DB 1; Length 407;
Best Local Similarity  35.6%;  Pred. No. 8.2e-05;
```

Matches 73; Conservative 11; Mismatches 52; Indels 69; Gaps 15;

```
Qy    3 GFG---VGVGGIPGVAG-VPGVGGVPGVGG-VPGVGISPEAQAAAAAKAKYGVGTPAAA 57
        |||   !|!:! |!.||:|| !!:|| ||!: |                      !:|
Db  213 GFGNQQPGSFGNQPGMGGRQPGFGNQPGMGGRQPGFGNQP---------------GMG----- 254

Qy   58 AAKAAAKKAAQFGLVPGVGVAPGV-GLAPGVGVAPGV-GVAPGVGVVAP 114
        |!:!:||!||!||!|:|!||!|||.||!||.|!|::| |!!:! |
Db  255 ----------GRQPGFGNQPGVGGRQPGFGNQPGMGGRQPGFGNQPGFGNQP 303

Qy  115 AIGPPEAQAAAAAKAKYGVGTPAAAAAAKAAQFGLVPGVGVAPGV-GVAPGV-GVA 172
        |||                          !!: | ||!:! ||:! |
Db  304 GMGGQQP----------------------GVG-----GRQPGFGNQPGMGGQQPGMGGQQ 336

Qy  173 PGV-GLAPGV-GVAPGV-GVAPGVG 194
        |!: | ||!:! |!:|
Db  337 PGMGGRQPGVGGRQPGMGGQQPGMG 361
```

RESULT 14

SPD2_NEPCL
ID SPD2_NEPCL　　　　STANDARD;　　　　PRT;　　　627 AA.
AC P46804;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPIDROIN 2 (DRAGLINE SILK FIBROIN 2) (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneoidea; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92405876; PubMed=1527052;
RA Hinman M.B., Lewis R.V.;
RT "Isolation of a clone encoding a second dragline silk fibroin.
RT Nephila clavipes dragline silk is a two-protein fiber.";
RL J. Biol. Chem. 267:19320-19324(1992).
CC -!- FUNCTION: THIS SPIDER MAJOR AMPULLATE SILK POSSESSES UNIQUE
CC   CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF
CC   PSEUDOCRISTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED
CC   WITH ELASTIC AMORPHOUS SEGMENTS.
CC -!- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 1, OF THE DRAGLINE SILK.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; M92913; AAA29381.1; -.
DR HSSP; P04002; 1WFA.
KW Silk; Repeat.
FT NON_TER   1   1
FT DOMAIN   1   530   15 APPROXIMATE TANDEM REPEATS.
FT REPEAT   1   36   1.
FT REPEAT   37   79   2.
FT REPEAT   80   121   3.
FT REPEAT   122   172   4.
FT REPEAT   173   213   5.
FT REPEAT   214   252   6.
FT REPEAT   253   283   7.
FT REPEAT   284   317   8.
FT REPEAT   318   359   9.
FT REPEAT   360   391   10.
FT REPEAT   392   428   11.
FT REPEAT   429   464   12.
FT REPEAT   465   488   13.
FT REPEAT   489   515   14.
FT REPEAT   516   530   15.
```

SQ SEQUENCE   627 AA;   54184 MW;   CB9B6377B9B2C594B CRC64;

```
Query Match             18.9%;  Score 192.5;  DB 1;  Length 627;
Best Local Similarity   36.5%;  Pred  No 0.00016;
Matches  84; Conservative  5; Mismatches 88; Indels 53; Gaps 14;
```

```
Qy    7 GVGGI-PGVAGVPGVGGVPGVGGVPGVGISPEAQAAAAAKAAKYGVG----------TPA 55
        ||!!! |                  !|!|||| ||!!||||||!||
Db  219 GPGSYGPGQQGPGGYG.-PGQQGLSGFG--SAAAAAAGPGQQGPGGYGPGQQGFSGPG 273

Qy   56 AAAKAAKKAAQPGLVPGV----GVAPGVGVGVAPGVGVGLA-PGV----GVAEGV-- 104
        :!!|||!|| |!|!||      :|!|||!|!||:|!|| |||:||
Db  274 SAAAAAAAAGPGGYGPGQQGPGGYGPGQQPPEAQAAAAAGPGQQGLGGYGPGQQG 333

Qy  105 -GVAPGVGVAPAIGPPEAQAAAAAKA-----AKYGVGT------PAAAAKAAAAQF 151
        :|!|||!|!|!                   ::|!|||         |!!|!||!||| ||
Db  334 PGGYGPGQQGPGGYGPGSASAAAAAGPGQQGPGGYGPGQQGPGGYGPGSASAAAAAGP 391

Qy  152 GLVPGVGVAPGVVAPGVGVAPGVGLAPGVGVGVAPGVGVAPGVGVAPGVGVAPAIGF 201
        |!|!||  | |  | | |||||!|||||:|  ||||!|||||:||||||:|| ||
Db  392 --------GPG-GYGPG-QQGPGG-GYAPGQQGPGSGPGSASAAAAAAGP 430
```

RESULT 15

CA18_MOUSE
ID CA18_MOUSE　　　　STANDARD;　　　　PRT;　　　743 AA.
AC Q00780;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE COLLAGEN ALPHA 1(VIII) CHAIN PRECURSOR.
GN COL8A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=92362626; PubMed=1499564;
RA Muragaki Y., Shiota C., Inoue M., Ooshima A., Olsen B.R.,
RA Ninomiya Y.;
RT "Alpha 1(VIII)-collagen gene transcripts encode a short-chain
RT collagen polypeptide and are expressed by various epithelial,
RT endothelial and mesenchymal cells in newborn mouse tissues.";
RL Eur. J. Biochem. 207:895-902(1992).
CC -!- FUNCTION: MAJOR COMPONENT OF THE DESCEMET'S MEMBRANE (BASEMENT
CC   MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
CC -!- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
CC   WITH ALPHA 2(VIII) TYPE COLLAGENS.
CC -!- TISSUE SPECIFICITY: HIGH LEVELS IN CALVARIUM, EYE & SKIN OF
CC   NEWBORN MICE; ALSO IN VARIOUS EPITHELIAL, ENDOTHELIAL AND
CC   MESENCHYMAL CELLS.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: STRONG, TO ALPHA 2 TYPES VIII AND X COLLAGENS.
CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; X66976; CAA47387.1; -.
DR EMBL; X66977; CAA47387.1; JOINED.
DR PIR; S23779; S23779.
DR MGD; MGI:88463; Col8a1.
DR InterPro; IPR000087; -.
DR InterPro; IPR001073; -.
```

```
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 7.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   PROSITE; PS01113; C1Q; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Cell adhesion; Collagen; Signal.
FT   SIGNAL        1     28       BY SIMILARITY.
FT   CHAIN        29    743       COLLAGEN ALPHA 1(VIII) CHAIN.
FT   DOMAIN       29    117       NONHELICAL REGION (NC2).
FT   DOMAIN      118    571       TRIPLE-HELICAL REGION (COL1).
FT   DOMAIN      572    743       NONHELICAL REGION (NC1).
FT   DOMAIN      608    743       C1Q.
SQ   SEQUENCE   743 AA;  73595 MW;  350E29502653EE0B CRC64;

Query Match
Best Local Similarity    18.5%;   Score 188;  DB 1;  Length 743;
                         33.5%;   Pred. No. 0.00032;
Matches  77;  Conservative  14;  Mismatches  89;  Indels  50;  Gaps  14;

Qy      6 VGVGGIPGVAGVPGVGGVPGCVPGVGGVPGVGISPEAQAAAAAKAAKYGVGTPAAAAAKAAAKA  65
          :|  ||||| | |||| | |:|||::|  |: |   : :|: :| | ||:|||||||| || ||
Db    187 IGPMGIPGPGPGPGPHGLPGIGKPGGPGL-PGQPGAKGERGPKGPPGPPGLQGPKGEKGF  245

Qy     66 AQFGLVPGVGVAPGVGVAPGVGVAPGVGVLAPGVGVAPG-----------VG  111
          :| ||: || |:||||| ||| ||| |:| |||| |:||    :|
Db    246 GMLGL-PGLKGPPGSMHGPPGPVGLPGVG-KPGVTGFPGPQGPLG-KPGPPGEPGPQGLIG  302

Qy    112 VAPAIGPPEAQAAAAAKAAKYGVGTPAAAAAKA----AAKAAQFGL--VPGVGVAPGVG  164
          |   ||      |||||||||||||| ||||||    : |  ||    ||||
Db    303 VPGVQGPP--------GMPGVGKPGQDGSRGQ2GFPGGKGEQ-GLPGLPGPPGLPGVG  351

Qy    165 VA--PGVGVAPGVGLAPGV----------GVAPGVGVAPGVGVAPAI-GP  201
          ||   :| |||    |||  : ||||    :|  | |||| |  : |:||
Db    352 KPGFPGKPGHRGIGGVPGVLGPRGEKGPIG-APGMGPPGPGEPGLPGIPGP  400
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001. 16:40:24 ; Search time 125.5 Seconds
                            (without alignments)
                            187.719 Million cell updates/sec

Title: US-09-340-736-2
Perfect score: 1017
Sequence: 1 FPGFGVGVGGIPGVAGVPGV.......GVAPGVGVAPGVGVAPAIGP 201

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs. 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
               Maximum Match 100%
               Listing first 45 summaries

Database :   SPTREMBL_15:*
            1: sp_archea:*
            2: sp_bacteria:*
            3: sp_fungi:*
            4: sp_human:*
            5: sp_invertebrate:*
            6: sp_mammal:*
            7: sp_mhc:*
            8: sp_organelle:*
            9: sp_phage:*
          10: sp_plant:*
         11: sp_rodent:*
         12: sp_unclassified:*
         13: sp_vertebrate:*
         14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	709.5	69.8	724	4	Q14233	Q14233 homo sapien
2	700	68.8	757	4	Q14234	Q14234 homo sapien
3	655.5	64.5	602	4	O15337	O15337 homo sapien
4	654.5	64.4	635	4	O15336	O15336 homo sapien
5	606	59.6	687	4	Q14235	Q14235 homo sapien
6	498	49.0	650	6	Q28099	Q28099 bos taurus
7	490	48.2	679	6	Q28097	Q28097 bos taurus
8	490	48.2	707	6	Q28098	Q28098 bos taurus
9	487	47.9	666	6	Q28096	Q28096 bos taurus
10	371.5	36.5	258	4	Q9DMP5	Q9DMP5 homo sapien
11	346.5	34.1	907	5	Q26675	Q26675 theileria a
12	291.5	28.7	1729	5	Q90617	Q90617 drosophila
13	287.5	28.3	1039	5	Q9VTR6	Q9VTR6 drosophila
14	260	25.6	988	6	O17434	O17434 nephila cla
15	249.5	24.5	76	6	Q28100	Q28100 bos taurus
16	240.5	23.6	120	14	Q88914	Q88914 thermoprote
17	240	23.6	1334	2	Q9KKR9	Q9KKR9 streptomyce
18	235.5	23.2	117	14	Q88913	Q88913 thermoprote
19	230	22.6	1489	2	O53559	O53559 mycobacteri
20	223.5	22.0	360	5	O16985	O16985 araneus dia
21	214	21.0	301	3	O01927	O01927 phytophthor
22	212	20.8	343	2	O01914	O01914 phytophthor
23	206	20.3	1567	11	Q93104	Q93104 mus musculu
24	205.5	20.2	1660	2	O53215	O53215 mycobacteri
25	204	20.1	738	5	O02402	O02402 plectada fu
26	203.5	20.0	347	3	O01916	O01916 phytophthor
27	200.5	19.7	1787	10	Q9MAX9	Q9MAX9 chlamydomon
28	195.5	19.2	219	2	O95589	O95589 mycobacteri
29	195.5	19.2	212	10	O53018	O53018 arabidopsis
30	192	18.9	2639	5	O76786	O76786 antheraea p
31	189.5	18.6	410	5	O16988	O16988 araneus dia
32	189.5	18.6	979	2	O01011	O01011 streptomyce
33	189	18.6	871	5	O44358	O44358 nephila cla
34	189	18.6	2249	5	Q9NFW4	Q9NFW4 nephila cla
35	188.5	18.5	610	5	Q9V5V8	Q9V5V8 drosophila
36	187.5	18.4	544	5	O46171	O46171 nephila cla
37	187.5	18.4	1306	2	O53775	O53775 mycobacteri
38	187.5	18.4	1802	5	O17163	O17163 brugia mala
39	186.5	18.3	767	2	O53435	O53435 mycobacteri
40	186	18.3	394	14	Q9J210	Q9J210 macaca muia
41	186	18.3	784	2	Q9RKL5	Q9RKL5 streptomyce
42	186	18.3	1884	5	Q9NFW2	Q9NFW2 nephila mad
43	185	18.2	416	5	O17374	O17374 caenorhabdi
44	183.5	18.0	413	2	P95236	P95236 mycobacteri
45	180.5	17.7	408	10	Q43698	Q43698 vigna ungui

ALIGNMENTS

RESULT 1
Q14233
ID   Q14233; Q14238;
AC   Q14233; Q14238;
DT   01-NOV-1996 (TrEMBLrel. 01; Created)
DT   01-NOV-1996 (TrEMBLrel. 01; Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15; Last annotation update)
DE   ELASTIN.
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87274906; PubMed=3038460;
RA   Indik Z., Yoon K., Morrow S.D., Cicila G., Rosenbloom J.,
RA   Rosenbloom J., Ornstein-Goldstein N.;
RT   "Structure of the 3' region of the human elastin gene: great abundance
RT   of Alu repetitive sequences and few coding sequences.";
RL   Connect. Tissue Res. 16:197-211(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87289668; PubMed=3039501;
RA   Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT   "Alternative splicing of human elastin mRNA indicated by sequence
RT   analysis of cloned genomic and complementary DNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
RN   [3]
RP   SEQUENCE OF 164-724 FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE=88156138; PubMed=2831431;
RA   Fazio M.J., Olsen D.R., Kuivaniemi H., Chu M.L., Davidson J.M.,
RA   Rosenbloom J., Uitto J.;
RT   "Isolation and characterization of human elastin cDNAs, and age-
RT   associated variation in elastin gene expression in cultured skin
RT   fibroblasts.";
RL   Lab. Invest. 58:270-277(1988).
DR   EMBL; M17282; AAC98394.1; -.
DR   EMBL; M16983; AAC98394.1; JOINED.
DR   EMBL; M17265; AAC98394.1; JOINED.

PRELIMINARY;   PRT;   724 AA.

DR   EMBL; M17266; AAC98394.1; JOINED.
DR   EMBL; M17267; AAC98394.1; JOINED.
DR   EMBL; M17268; AAC98394.1; JOINED.
DR   EMBL; M17270; AAC98394.1; JOINED.
DR   EMBL; M17271; AAC98394.1; JOINED.
DR   EMBL; M17272; AAC98394.1; JOINED.
DR   EMBL; M17273; AAC98394.1; JOINED.
DR   EMBL; M17275; AAC98394.1; JOINED.
DR   EMBL; M17276; AAC98394.1; JOINED.
DR   EMBL; M17277; AAC98394.1; JOINED.
DR   EMBL; M17278; AAC98394.1; JOINED.
DR   EMBL; M17279; AAC98394.1; JOINED.
DR   EMBL; M17280; AAC98394.1; JOINED.
DR   EMBL; M17281; AAC98394.1; JOINED.
DR   EMBL; M24782; AAA53190.1; -.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
SQ   SEQUENCE   724 AA; 62664 MW; 611558622BEDF3D CRC64;

Query Match   69.8%; Score 709.5; DB 4; Length 724;
Best Local Similarity   53.8%; Pred. No. 1.3e-37;
Matches 164; Conservative   6; Mismatches   26; Indels 109; Gaps   6;

Qy   1 FPGFGVGVGGIPGVAGVPGVGVPGVSVGVPGVGISPEAQAAAAKAAKYGVGTPAAAAK 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   404 FPGFGVGVGGIPGVAGVPSVGGVPGVGISPEAQAAAAKAAKYGVGTPAAAAK 463

Qy   61 AAAKAAQPGLVPGVGVAPGVGVAPGVPCVGLAPGVGVAPGVGVAPGVGVAPAIGPPE 120
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   464 AAAKAAQPGLVPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGIGPGG 523

Qy   121 AQAA-- 124
          ||
Db   524 VAAAAKSAAKYAAKAQLRAAAGIPGLGVGVGVPGSLGVGVGVPGFGAV 583

Qy   125 ---AAAAAKYGVGTP-----------------------AAAAAKAAAKAAQPGLVP 155
          ||||||| ||||||
Db   584 FGALAAAAAAKYGAAVPGVLGGLGALGGVGIPGGVVGAGPAAAAAKAAAKAAQPGLVG 643

Qy   156 GVGVAPGVGV----APGVGVAPGV------------------GLAPGVVAP--GVGVAPG 192
          |||| |||| ||||| |||||||||
Db   644 AAGLG-GLGTVGGLGTVPGVGGLGGIPPAAAKAAKYGAAGLGGVLGGAGQEPLGGVAAPFG 702

Qy   193 VGVAP 197
          |||||
Db   703 FGLSP 707

LI   2
ID   Q14234   PRELIMINARY;   PRT; 757 AA.
AC   Q14234;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN.
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87274906; PubMed=3038460;
RA   Indik Z., Yoon K., Morrow S.D., Cicila G., Rosenbloom J.,
RA   Rosenbloom J. Ornstein-Goldstein N.;
RT   "Structure of the 3' region of the human elastin gene: great abundance
RT   of Alu repetitive sequences and few coding sequences.";
RL   Connect. Tissue Res. 16:197-211(1987).
RN   [2]

RP   SEQUENCE FROM N.A.
RX   MEDLINE=87289668; PubMed=3039501;
RA   Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT   "Alternative splicing of human elastin mRNA indicated by sequence
RT   analysis of cloned genomic and complementary DNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
DR   EMBL; M17282; AAC98395.1; -.
DR   EMBL; M16983; AAC98395.1; JOINED.
DR   EMBL; M17265; AAC98395.1; JOINED.
DR   EMBL; M17266; AAC98395.1; JOINED.
DR   EMBL; M17267; AAC98395.1; JOINED.
DR   EMBL; M17268; AAC98395.1; JOINED.
DR   EMBL; M17270; AAC98395.1; JOINED.
DR   EMBL; M17271; AAC98395.1; JOINED.
DR   EMBL; M17272; AAC98395.1; JOINED.
DR   EMBL; M17273; AAC98395.1; JOINED.
DR   EMBL; M17274; AAC98395.1; JOINED.
DR   EMBL; M17275; AAC98395.1; JOINED.
DR   EMBL; M17276; AAC98395.1; JOINED.
DR   EMBL; M17277; AAC98395.1; JOINED.
DR   EMBL; M17278; AAC98395.1; JOINED.
DR   EMBL; M17279; AAC98395.1; JOINED.
DR   EMBL; M17280; AAC98395.1; JOINED.
DR   EMBL; M17281; AAC98395.1; JOINED.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
SQ   SEQUENCE   757 AA; 66236 MW; 23B7FE5B8AF85CA8 CRC64;

Query Match   68.8%; Score 700; DB 4; Length 757;
Best Local Similarity   76.1%; Pred. No. 5.4e-37;
Matches 156; Conservative   7; Mismatches   24; Indels   18; Gaps   8;

Qy   1 FPGFGVGVGGIPGVAGVPGVGVPGVGVPGVGISPEAQAAAAKAAKYGVGTPAAAAAK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   404 FPGFGVGVGGIPGVAGVPSVGGVPGVGVPGVGISPEAQAAAAKAAKYGVGTPAAAAAK 463

Qy   61 AAAKAAQPGLVPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPAIGPPE 120
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   464 AAAKAAQPGLVPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGIGP- 521

Qy   121 AQAAAAAKAAKYGVGTPAAAAKAAAKAAAKAAQFGL---VPGVGVAPGVGVAPGV-G 176
          |||:|:| ::|:|| :||:|| || ||||||| || || |
Db   522 GGVAAAAKSA------AKVAAKAQLRAAA-GLGAGIPGLGV--GVGV-PGLGVGAGVPG 570

Qy   177 LAPGVGVAPGVGVAPGVGVAPAIGP 201
          |||:||
Db   571 LGVGAGV-PGFGAGADEGVRRSLSP 594

RESULT   3
ID   O15337   PRELIMINARY;   PRT; 602 AA.
AC   O15337;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Li D.Y., Toland A.E., Boak B.B., Atkinson D.L., Ensing G.J.,
RA   Morris C.A., Keating M.T.;
RL   Hum. Mol. Genet. 0:0-0(1997).
DR   EMBL; U93037; AAB65620.1; -.
DR   EMBL; U93034; AAB65620.1; JOINED.
DR   EMBL; U93035; AAB65620.1; JOINED.

```
DR   EMBL; U93036; AAB65620.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   INTERPRO; IPR002185; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00569; DOPAMINED4R.
DR   PRINTS; PR00762; CLCHANNEL.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR   PROSITE; PS004453; FKBP_PPIASE_1; UNKNOWN_1.
FT   NON_TER         1       1
FT   NON_TER       602     602
SQ   SEQUENCE   602 AA;  51807 MW;  53B5B9A71BF04807 CRC64;

Query Match        64.5%;  Score 655.5; DB 4;  Length 602;
Best Local Similarity  65.5%;  Pred. No. 2.7e-34;
Matches 154;  Conservative   6;  Mismatches  18,  Indels  57;  Gaps  10;

Qy     1 FPGFGVGVGGIPGVAGVPGVGVGGVPGVGVGGVPGVGISPEAQAAAARKAAKY----------- 49
         |||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   376 FPGFGVGVGGIPGVAGVPGVGVGGVPGVGCVPGVGISPEAQAAAARAAKYGAAGAGVLGGL 435

Qy    50 -----------------GVGTPAAAAAAKAKAAQP------GLVPGVGVAPGVGVAPG 85
                          |||||||||||||||      |||||||||||||||
Db   436 VPGADPGAVPGVPGVPGTGCVPGVAPGVGTPAAAAAKAKAAQFALLNLAGLVPGVGVAPGVGVAPG 495

Qy    86 VGVAPGVGLAPGVGVAPGVGVAPGVGVAPGV-GLAPGVGVAPGVGVAPGVGVAPAAAKAA 145
         ||||||||||||||||||||||||       |||||||||||||||
Db   496 VGVAPGVGLAPGVGVAPGVGVAPGIGP--GGVAAAAKSA------AKVAAKAQ 546

Qy   146 AKAAQFGL---VPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGVGVAP 197
         |||||||   :|||| |||||  |||||||| ||
Db   547 LRAAA-GLGAGIPGLGV--GVGV-PGLGVGAGV--PGFRAVP 591

RESULT 4
O15336
ID   O15336;       PRELIMINARY;      PRT;       635 AA.
AC   O15336;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Li D.Y., Toland A.E., Boak B.B., Atkinson D.L., Ensing G.J.,
RA   Morris C.A., Keating M.T.;
RL   Hum. Mol. Genet. 0:0-0(1997).
DR   EMBL; U93037; AAB65621.1; -.
DR   EMBL; U93034; AAB65621.1; JOINED.
DR   EMBL; U93035; AAB65621.1; JOINED.
DR   EMBL; U93036; AAB65621.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   INTERPRO; IPR001807; -.
DR   INTERPRO; IPR002185; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00569; DOPAMINED4R.
DR   PRINTS; PR00762; CLCHANNEL.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR   PROSITE; PS004453; FKBP_PPIASE_1; UNKNOWN_1.
FT   NON_TER         1       1
FT   NON_TER       635     635
SQ   SEQUENCE   635 AA;  55279 MW;  72950C364127B2A4 CRC64;
```

```
Query Match        64.4%;  Score 654.5; DB 4;  Length 635;
Best Local Similarity  64.6%;  Pred. No. 3.3e-34;
Matches 155;  Conservative   7;  Mismatches  25,  Indels  53;  Gaps  10;

Qy     1 FPGFGVGVGGIPGVAGVPGVGVGGVPGVGVGGVPGVGISPEAQAAAARKAAKY----------- 49
         |||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   376 FPGFGVGVGGIPGVAGVPGVGVGGVPGVGCVPGVGISPEAQAAAARAAKYGAAGAGVLGGL 435

Qy    50 -----------------GVGTPAAAAAAKAKAAQF------GLVPGVGVAPGVGVAPG 85
                          |||||||||||||||      |||||||||||||||
Db   436 VPGADPGAVPGVPGVPGTGCVPGVAPGVGTPAAAAAKAKAAQFALLNLAGLVPGVGVAPGVGVAPG 495

Qy    86 VGVAPGVGLAPGVGVAPGVGVAPGVGVAPGV-GLAPGVGVAPGVGVAPGVGVAPAAAKAA 145
         ||||||||||||||||||||||||       |||||||||||||||
Db   496 VGVAPGVGLAPGVGVAPGVGVAPGIGP--GGVAAAAKSA------AKVAAKAQ 546

Qy   146 AKAAQFGL--VPGVGVAPGVGVAPCSVGVAPGV-GLAPGVGYAPGVGVAPGVGVAPAIGP 201
         |||||||  :||||| |||::||  ||||||||| ||  :|:|
Db   547 LRAAA-GLGAGIPGLGV--GVGV-FSLGVGAGVPGLGVGAGV-PGFRAGADEGYRRSLSP 601

RESULT 5
Q14235
ID   Q14235;       PRELIMINARY;      PRT;       687 AA.
AC   Q14235;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN.
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87274906; PubMed=3038460;
RA   Indik Z., Yoon K., Morrow S.D., Cicila G., Rosenbloom J.,
RA   Rosenbloom J., Ornstein-Goldstein N.;
RT   "Structure of the 3' region of the human elastin gene: great abundance
RT   of Alu repetitive sequences and few coding sequences.";
RL   Connect. Tissue Res. 16:197-211(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87289658; PubMed=3039501;
RA   Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT   "Alternative splicing of human elastin mRNA indicated by sequence
RT   analysis of cloned genomic and complementary DNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
DR   EMBL; M17283; AAC98393.1; -.
DR   EMBL; M16983; AAC98393.1; JOINED.
DR   EMBL; M17265; AAC98393.1; JOINED.
DR   EMBL; M17266; AAC98393.1; JOINED.
DR   EMBL; M17267; AAC98393.1; JOINED.
DR   EMBL; M17268; AAC98393.1; JOINED.
DR   EMBL; M17271; AAC98393.1; JOINED.
DR   EMBL; M17272; AAC98393.1; JOINED.
DR   EMBL; M17273; AAC98393.1; JOINED.
DR   EMBL; M17275; AAC98393.1; JOINED.
DR   EMBL; M17276; AAC98393.1; JOINED.
DR   EMBL; M17277; AAC98393.1; JOINED.
DR   EMBL; M17278; AAC98393.1; JOINED.
DR   EMBL; M17279; AAC98393.1; JOINED.
DR   EMBL; M17281; AAC98393.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR   PROSITE; PS004453; FKBP_PPIASE_1; UNKNOWN_1.
SQ   SEQUENCE   687 AA;  59579 MW;  9D5AC3C4D9F9B98E CRC64;

Query Match        59.6%;  Score 606;  DB 4;  Length 687;
```

```
Best Local Similarity  50.0%;  Pred. No. 3.9e-31;
Matches 146;  Conservative 6;  Mismatches 20;  Indels 120;  Gaps 8;

Qy    1 FPGFGVGVGGIPGVAGVPGVGGVPGVGGVGISPEAQAAAAKAAKYGVGTPAAAAAK    60
          |||||||||||||||||||||||||||||||||||||||
Db  404 FPGFGVGVGGIPGVAGVPSVGGVPGGVPGVGGISPEAQAAAAAKAAKY------------   452

Qy   61 AAAKAAQPGLVPGVGVGARPGVGVGVAPGVGVAPGVGVAPGVGVAPAIGPPE           120
          |||||||||||||||||||||||||
Db  453 --------GLVPGVGVGARPGVGVGARPGVGLARPGVGVGVAPGVGVAPGVGPGG        504

Qy  121 AQPA----------                                                 124
          ||
Db  505 VAAAAKSRAKVAAKAQLRAAAGLGAGIPGLGVGVGVPGLGVGVGVPGLGVGAGVPGGFGAV  564

Qy  125 ----AAAKAAKYGVGTP----------                                    155
             ||||||||||
Db  565 PGALAAAKAAAKYGAAVPGVLGGLGALGGVGYPGGVVGAGPAAAAAAAAAKAAQFGLVG    624

Qy  156 GVGVAPGVGVAPGVGVLAPGVGVAP--------GVGVAPGVGVAP                  197
          ||    ||| ||  || |||    ||        || ||
Db  625 AAGLG-GLGVG-GLGV-PGVG---GLGGIPAAAAAKAAKYGVAARPGPGLSP           670


RESULT 6
Q28099
ID   Q28099;
AC   Q28099;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN-CBEI1 (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992576;
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J.;
RT   "Structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C, Cicila G., Yoon K. Rosenbloom J.;
RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30499.1; JOINED.
DR   EMBL; M11422; AAA30499.1; JOINED.
DR   EMBL; M19366; AAA30499.1; JOINED.
DR   EMBL; M19368; AAA30499.1; JOINED.
DR   EMBL; M19369; AAA30499.1; JOINED.
DR   EMBL; M19367; AAA30499.1; JOINED.
DR   EMBL; M19371; AAA30499.1; JOINED.
DR   EMBL; M22771; AAA30499.1; JOINED.
DR   EMBL; M22772; AAA30499.1; JOINED.
DR   EMBL; M22773; AAA30499.1; JOINED.
DR   EMBL; M22774; AAA30499.1; JOINED.
DR   EMBL; M22775; AAA30499.1; JOINED.
DR   EMBL; M22988; AAA30499.1; JOINED.
DR   EMBL; M23010; AAA30499.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
FT   NON_TER       1      1
SQ   SEQUENCE   650 AA;  55373 MW;  CD21ABB3E9076AD7 CRC64;


Query Match   49.0%;  Score 498;  DB 5;  Length 650;
```

```
Best Local Similarity  42.4%;  Pred. No. 2.2e-24;
Matches 145;  Conservative 12;  Mismatches 35;  Indels 150;  Gaps 16;

Qy    1 FPGFGVGVGGIPGVAGV---PGVG-----GVPGVGGY----PGV--------GISP      36
          ||||  ||:    :|: |||   ||:  |||| |||        |||            |:|
Db  115 FPGIGV-LPGVPTGAGVKFKAPGGGGAPAGIPSVGFPGGQQPGVPLGYPIKAPKLPGVGP   174

Qy   37 EAQAAAAAKAAKY--------GVGTP-AAAAKAIAA                           63
          :| |||| |||             ||| |||||||
Db  175 QAAAAAKAAAKIGAGGAGVLPGVGVGVGGAGIPGAPGAIPGIGGIAGVGAPDAAAAAAAA   234

Qy   64 KAAQPGL---------VPGVGVAPGVGVAPGVGL----APGVGV---APGVGV         106
          ||:||          ||||| ||||| ||||         |||||    |||||
Db  235 KAAKPGAAGGFPGVGVPGVGV-PGVGV-PGVGVGVPGVGVGVPGVGVGVPGVGVGVP      291

Qy  107 ---APGVGVAPIGP---------PEAQA                                   123
          |:|                      |:|
Db  292 GVGVPGVGVPGAVSPAAAAKAAAKAAKPGARGGVGIGGIPTGVGPGGPPGIGDAAAAQA   351

Qy  124 AAAAKAAKY-------------GVGIPAAAAAKAAKAAQFGLV                    154
          ||||||                 |||||||||||||||||||
Db  352 AAAAKAAKIGAGGVAGLGGLVFGAIPGVPGVGGYPGVGGVGIPARAAAAKARAAQPGLG   411

Qy  155 PGVGAPGVGVAPGVGVGVAPGVGLAPGVGVAPGVGVGVA                        196
          ||||| ||||| |||| ||||||| |||| :||: || :|
Db  412 PGVGVAPGVGVVPGVGVVFGVGVVPGGIGLGPGGSVIGAGYPAA                   453


RESULT 7
Q28097
ID   Q28097;
AC   Q28097;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN-CBEI2 (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992576;
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J.;
RT   "Structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Cicila G., Yoon K., Rosenbloom J.;
RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30500.1; JOINED.
DR   EMBL; M11422; AAA30500.1; JOINED.
DR   EMBL; M19366; AAA30500.1; JOINED.
DR   EMBL; M19369; AAA30500.1; JOINED.
DR   EMBL; M19367; AAA30500.1; JOINED.
DR   EMBL; M19358; AAA30500.1; JOINED.
DR   EMBL; M19369; AAA30500.1; JOINED.
DR   EMBL; M19370; AAA30500.1; JOINED.
DR   EMBL; M19371; AAA30500.1; JOINED.
DR   EMBL; M22771; AAA30500.1; JOINED.
DR   EMBL; M22772; AAA30500.1; JOINED.
DR   EMBL; M22773; AAA30500.1; JOINED.
DR   EMBL; M22774; AAA30500.1; JOINED.
DR   EMBL; M22988; AAA30500.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   INTERPRO; IPR001459; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00959; MEVGALKINASE.
```

```
DR   EMBL; M22771; AAA30498.1; JOINED.
DR   EMBL; M22772; AAA30498.1; JOINED.
DR   EMBL; M22773; AAA30498.1; JOINED.
DR   EMBL; M22774; AAA30498.1; JOINED.
DR   EMBL; M22775; AAA30498.1; JOINED.
DR   EMBL; M29988; AAA30498.1; JOINED.
DR   EMBL; M23010; AAA30498.1; JOINED.
DR   INTERPRO; IPR001439; -.
DR   INTERPRO; IPR001104; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00959; MEVGALKINASE.
FT   NON_TER           1      1
SQ   SEQUENCE  707 AA;  60346 MW;  FDFD559BAB34CE33 CRC64;

Query Match              48.2%;  Score 490;  DB 6;  Length 707;
Best Local Similarity   38.8%;  Pred. No. 7.6e-24;
Matches 146;  Conservative 12;  Mismatches 34;  Indels 184;  Gaps 16;

Qy       1 PPGFGVGVGGIPGVAGV----PGVG-------GVP------------------- 30
           ||| || |:| ||| ||    |:||| ||         ||
Db     116 FPGIGV-LPGVFFGAGVKPKAPGGGAFAGIPGVGVPFGGQQPGVPLGYPIKAPKLPAGYG 174

Qy      31 ----------------------GVGISPERQAAAARAAKY------------------ 49
                                 | |: |:| |||| |||
Db     175 LPYKTGKLPYGFGPGVAGSAGKAGYFTGTGVGPQAAAAARAAAKLGAGGAGVLPGVGV 234

Qy      50 ---------GVGTP-AAAARAAKAAQFGL--------VPGVGVAPGVG 81
                    ||:| ||||| ||||| ||:||             |||||| ||||
Db     235 GGAGIPGAPGAIPGIGGIAGVGAPDAAAAAAAAKAAKFGAAGGFPGVGVPGVGV-PGVG 293

Qy      82 VAPGVGVAPGVGL-----APGVGV-----APGVGVAPAIGP------------- 118
            | ||||| |||:  ||||   |||||   |:|
Db     294 V-PGVGV-PSVGVPGVGVPGVGVPGVGVPGVGVPGAVSPAAAAKAAAKAA 351

Qy     119 -----------------PERQAAAAKAAKY------------------ 132
                             ||||||||||
Db     352 KFGARGGVGIGGIPFPGVGPGGFPGIGDAAAAQAAAAAKAAKIGAGGVGALGGLVPGAPG 411

Qy     133 -----------GVGTPAAAAAKAAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPG 180
                      ||||||||||||||| ||||||||||| ||||||||| |:|
Db     412 AIPGVPGVGVGVAPGVGIPAAAAAKAAAKAAGFGLGPGVGVAPGVGVVPGVGVVPGSVGVAPG 471

Qy     181 VGVAPGVGVAPGVGVA 196
            |:|| || || |
Db     472 IGLGPGGVIGAGVPAA 487

RESULT 9
ID   Q28096        PRELIMINARY;      PRT;   656 AA.
AC   Q28096;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN-CBEL3 (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992576;
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J., Yoon K.;
RT   "Structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Cicila G., Yoon K., Rosenbloom J.;
```

```
FT   NON_TER           1      1
SQ   SEQUENCE  679 AA;  57652 MW;  EB3C019F3BD761BD CRC64;

Query Match              48.2%;  Score 490;  DB 6;  Length 679;
Best Local Similarity   38.8%;  Pred. No. 7.3e-24;
Matches 146;  Conservative 12;  Mismatches 34;  Indels 184;  Gaps 16;

Qy       1 PPGFGVGVGGIPGVAGV----PGVG-------GVP------------------- 30
           ||| || |:| ||| ||    |:||| ||         ||
Db     116 FPGIGV-LPGVFFGAGVKPKAPGGGAFAGIPGVGVPFGGQQPGVPLGYPIKAPKLPAGYG 174

Qy      31 ----------------------GVGISPERQAAAARAAKY------------------ 49
                                 | |: |:| |||| |||
Db     175 LPYKTGKLPYGFGPGVAGSAGKAGYFTGTGVGPQAAAAARAAAKLGAGGAGVLPGVGV 234

Qy      50 ---------GVGTP-AAAARAAKAAQFGL--------VPGVGVAPGVG 81
                    ||:| ||||| ||||| ||:||             |||||| ||||
Db     235 GGAGIPGAPGAIPGIGGIAGVGAPDAAAAAAAAKAAKFGAAGGFPGVGVPGVGV-PGVG 293

Qy      82 VAPGVGVAPGVGL-----APGVGV-----APGVGVAPAIGP------------- 118
            | ||||| |||:  ||||   |||||   |:|
Db     294 V-PGVGV-PSVGVPGVGVPGVGVPGVGVPGVGVPGAVSPAAAAKAAAKAA 351

Qy     119 -----------------PERQAAAAKAAKY------------------ 132
                             ||||||||||
Db     352 KFGARGGVGIGGIPFPGVGPGGFPGIGDAAAAQAAAAAKAAKIGAGGVGALGGLVPGAPG 411

Qy     133 -----------GVGTPAAAAAKAAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGLAPG 180
                      ||||||||||||||| ||||||||||| ||||||||| |:|
Db     412 AIPGVPGVGVGVAPGVGIPAAAAAKAAAKAAGFGLGPGVGVAPGVGVVPGVGVVPGSVGVAPG 471

Qy     181 VGVAPGVGVAPGVGVA 196
            |:|| || || |
Db     472 IGLGPGGVIGAGVPAA 487

RESULT 8
ID   Q28098        PRELIMINARY;      PRT;   707 AA.
AC   Q28098;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992576;
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J., Yoon K.;
RT   "Structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Cicila G., Yoon K., Rosenbloom J.;
RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30498.1; JOINED.
DR   EMBL; M11422; AAA30498.1; JOINED.
DR   EMBL; M19366; AAA30498.1; JOINED.
DR   EMBL; M19367; AAA30498.1; JOINED.
DR   EMBL; M19368; AAA30498.1; JOINED.
DR   EMBL; M19369; AAA30498.1; JOINED.
DR   EMBL; M19370; AAA30498.1; JOINED.
DR   EMBL; M19371; AAA30498.1; JOINED.
```

RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30501.1; JOINED.
DR   EMBL; M11422; AAA30501.1; JOINED.
DR   EMBL; M19366; AAA30501.1; JOINED.
DR   EMBL; M19367; AAA30501.1; JOINED.
DR   EMBL; M19368; AAA30501.1; JOINED.
DR   EMBL; M19369; AAA30501.1; JOINED.
DR   EMBL; M19370; AAA30501.1; JOINED.
DR   EMBL; M19371; AAA30501.1; JOINED.
DR   EMBL; M22771; AAA30501.1; JOINED.
DR   EMBL; M22772; AAA30501.1; JOINED.
DR   EMBL; M22773; AAA30501.1; JOINED.
DR   EMBL; M22774; AAA30501.1; JOINED.
DR   EMBL; M22775; AAA30501.1; JOINED.
DR   EMBL; M22988; AAA30501.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
FT   NON_TER   1
SQ   SEQUENCE   666 AA;   56435 MW;   BCB5E62632BE1B71 CRC64;

Query Match   47.9%; Score 487; DB 6; Length 666;
Best Local Similarity   41.4%; Pred. No. 1.1e-23;
Matches 145; Conservative 12; Mismatches 35; Indels 158; Gaps 17;

Qy   1 FPGFGVGVGGIPGVAGV----PGVG----GVPGVGGV-----PGV----------GISP 36
          |||| ||:|:|| || |||| || || |:| |:|
Db 116 FPGIGV-LPGVPGAGVKPKAPGGGGAFAGIPGVPGPGGQQPGVFLGTPIKAPKLPGVGP 174

Qy   37 EAQAAAAAAKY--------------------------GVGTP-AAAAAKAAA 63
          :| ||||| | ||||| | || ||| |||
Db 175 QAAAAANAAAKLGAGAGAVLPGVGVGGAGIPGAPGAIPGIGGIAGVGAPDAAAAAAAAA 234

Qy   64 KAAQFGL----------VPGVGVAPGVGVAPGVGVAPGVGL----APGVGV----APGVGV 106
          |||:|:| |||||| |||| ||| || |||||
Db 235 KAAKFGAAAGGFPGVGVPGVGV-PGVGV-PGVGVPGVGVPGVGVPGVGVP 291

Qy 107 --APGVGVAPAIGP------------------------------------PEAQA 123
          ||||| |:| |||
Db 292 GVGVPGVGVPGAVSPAAAAKAAAKAAKRAKFGARGGVGIGGTPTFGVGPGGPGIGDAAAACA 351

Qy 124 AAAAKAAKY--------------------------GVSTPAAAAAKAAAKAAQFGLV 154
          |||||||| ||| ||||||||||||||||||
Db 352 AAAAKAKAGGAGGVGALGGLVPGAPGAIPGVGPGVGVGVPGVGIPAAAAKAAAKAAQFGLA 411

Qy 155 -----PGVGVAPGVGVAPGVGLAPGVGLAPGVGVAPGVGVAPGVGVA 196
          ||||||||| ||||| || |||:|:|:|| :| || |
Db 412 LLAPAGLGPGVGVAPGVGVPGVGVPGVGVAPGIGLGPGGVIGAGVPAA 461

RESULT 10
Q9DMF5
ID   Q9DMF5   PRELIMINARY;   PRT;   258 AA.
AC   Q9DMF5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   ELASTIN (FRAGMENT).
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96411691; PubMed=8812460;
RA   Osborne L.R., Martindale D.W., Scherer S.W., Shi X.-M., Huizenga J.,
RA   Heng H.H.Q., Costa T., Pober B., Lew L., Brinkman J., Rommens J.,
RA   Koop B.F., Tsui L.-C.;
RT   "Identification of genes from a 500-kb region at 7q11.23 that is
RT   commonly deleted in Williams syndrome patients.";

RL   Genomics 36:328-336(1996).
DR   EMBL; U63721; AAC13884.1; -.
DR   INTERPRO; IPR001451; -.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
FT   NON_TER   1   1
SQ   SEQUENCE   258 AA;   21990 MW;   C39BF7298D0E05D2 CRC64;

Query Match   36.5%; Score 371.5; DB 4; Length 258;
Best Local Similarity   40.9%; Pred. No. 8e-17;
Matches 97; Conservative   6; Mismatches 25; Indels 109; Gaps   6;

Qy   69 GLVPGVGVAPGVGVAPGVGLAPGVGVGVAPGVGVAPGVGVAPAIGPPEAQAA---- 124
          ||||||||||||||||||||| |||||||||||||||||||| ||
Db   6 GLVPGVGVAPGVGVAPGVGLAPGVGVAPGVGVAPGIGPGGVAAAAKSA 65

Qy 125 --AAAK 128
                                                            ||||
Db   66 AKVAAKAQLRAAAGLGAGIPGLGTVGVGVPGLGVGAGVPGVPGGAVPGALAAAK 125

Qy 129 AAKYGVGTP--------------------AAAAKAAAKAAQFGLVPGVGVGVAPGV 163
          ||||| |||||||||||||||||||||| :| :|
Db 126 AAKYGAAVPGVGLGLGALGGVGIPGGVVGAGPAAAAAAKAAKAAQFGLVGAAGLG-GL 184

Qy 164 GV----APGVGAPGV------------------GLAPGVVAP--GVGVGVGVGVAP 197
          || || || ||:| :| || || | ||:|
Db 185 GVGGLGVPGVGVGLGGIPPAAAAKAAAKYGAAGLGGVLGGAGQPPLGGVAARPGFGLSP 241

RESULT 11
Q26675
ID   Q26675   PRELIMINARY;   PRT;   907 AA.
AC   Q26675;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   SPOROZOITE SURFACE ANTIGEN.
GN   SPAG-1.
OS   Theileria annulata.
OC   Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
OC   Theileria.
OX   NCBI_TaxID=5874;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92365719; PubMed=1501630;
RA   Hall R., Hunt P.D., Carrington M., Simmons D.L., Williamson S.,
RA   Tait A.;
RT   "Mimicry of elastin repetitive motifs by Theileria annulata sporozoite
RT   surface antigen.";
RL   Mol. Biochem. Parasitol. 53:105-112(1992).
DR   EMBL; M63017; AAA30134.1; -.
DR   INTERPRO; IPR000600; -.
DR   PROSITE; PS01125; ROK; UNKNOWN_1.
KW   Sporozoite.
SQ   SEQUENCE   907 AA;   91885 MW;   589CE55C740D4835 CRC64;

Query Match   34.1%; Score 346.5; DB 5; Length 907;
Best Local Similarity   39.1%; Pred. No. 9.7e-15;
Matches 111; Conservative 14; Mismatches 64; Indels 95; Gaps 16;

Qy   2 PGFGV-GVG----GIPGVAGVPGVG----GVPGVGGVPGVGISPEAQAAAAAKAAKYGVG 52
          || || ||| |||||||| |||||||:::| || |||
Db 179 PGVGVPGVGVPGVGVPGVGV-GVPGVGVGVAPGVGVAPGVGVAPGVGVPGVVAPGVGVG 237

Qy   53 T----PAAAAKAAAKAAQFGLVPGVGVVARGVGVA----------PGVGLAPGVGVAP 102
          | || ||:: | |||||| :|| :|||||| ||:| ||||||
Db 238 ADSSGLPGSGGLGAGAKAGK-QGGSGLQGPGGVGVPGVWPGVGVAASSSPGKPPGVGAGVNP 296

Qy 103 GVGV---------APGVGAPA--IGPPEAQ----------- 122
          ||:| ||:| :| |:|||||
Db 297 GVGVRAQQGVIIGAPGVAGVPGGKPGGPVSQELELKSDTEINESGSSSEGEDDDEEEEE 356

```
Qy   123 ------AAAAAKAAK YGVGTPAAAAAKAAAKAA----QPGLV---------------- 154
         || || || : ||   :|:||   :|: ||
Db   357 ENKSTSSKGAGGKAGK-GQGSVSPGGGSASQTSPTTTPQSGLASSGSHAQQSPQQDPAP 415

Qy   155 --PGVGVAPGVGVAPGVGVGLAPGVGVAPGVGVAPGVGVA 196
           || |||| ||||||||||| |||||||||||| ||||
Db   416 SKPSGGGVPGVGV-PGVGV-PGVGV VAPGVGVPGVGGA 456
```

RESULT 12
ID   Q90617
AC   Q90617;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PERICARDINE.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RC   SEQUENCE FROM N.A.
RC   STRAIN=OREGON R;
RA   Chartier A., Astier M., Zaffran S., Semeriva M., Gratecos D.;
RT   "Characterization of an extracellular matrix component specifically
RT   expressed in the basal lamina of the dorsal vessel in Drosophila
RT   melanogaster.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF203342; AAF13280.1; -.
SQ   SEQUENCE   1729 AA;   164661 MW;   DA9B1B7FA85CDEDB CRC64;

```
Query Match   28.7%;  Score 291.5;  DB 5;  Length 1729;
Best Local Similarity  35.4%;  Pred. No. 5e-11;
Matches   80;  Conservative   15;  Mismatches   92;  Indels   39;  Gaps   11;

Qy    2 PGFGV--GVGGIPGVAGVPGVGGVGVG---------GVPGVGISPEAQAAAAKAAKYGVGTP 54
          || ||  ||: || |    |||| |||        |||||| :|||   ||||||| ||
Db  296 PGYGTQPGIGGQTG-AGQPGYGSQP-AGQPGYGSQP--GIGGQTGAGQPGYGTQ 352

Qy   55 AAAAAKAAAKAAQPGLVPGVG-----VAPGVGVAPGVG-----VAPGVGLAPGVGVAPGV 104
          ||||||||| |||| |||| :    |   ||| |||       |||||| ||||| |
Db  353 PGIGGQTGAGQPGYGTQPGIG-----GQTGAGQPGYGTQPGIGGQTGAGQPGYGSQ 412

Qy  105 GVAPGVGVAPAIGPPEAQAAAAAAKYGVGTPAAAAAKAAAKAAQPGLVPGVG-----V 159
          ||| ||| || |||||||||||||     ||||||||||| ||||| |||| :    |
Db  413 G-QPGYGTQPGIG-----GGTGAGQPGYGTQPGIGAQTGAGQPGYGVQPGIGGQTGAG 464

Qy  160 APGVGVAPGVGLAPGVGVAPGVG-----APGVTAPAIG 200
          |||||||| |||| |||| |||      |||| || |
Db  465 QPGYGTQPGIGVTGAG-QPGYGSQPGIGAQTGAGQPGYGSQPGIG 509
```

RESULT 13
ID   Q9VTR6
AC   Q9VTR6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   CG5700 PROTEIN.
GN   CG5700.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merklilov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003543; AAF49980.1; -.
DR   FLYBASE; FBgn0036231; CG5700.
SQ   SEQUENCE   1039 AA;   102635 MW;   6A710871139BD01F CRC64;

```
Query Match   28.33%;  Score 287.5;  DB 5;  Length 1039;
Best Local Similarity  36.0%;  Pred. No. 5.5e-11;
Matches   81;  Conservative   14;  Mismatches   99;  Indels   31;  Gaps   10;

Qy    2 PGFG---VGVGGIPGVAGVPGVG---GVPGVGG------------VPGVGISPEAQAAAAAKAAAKYG 50
          ||||   ||: ||      |||   |||||||            |||||| :|||  || |
Db  203 PGFGGQPGIGGQTG-AGQPGVGTGQPGIGGQTGSGRQPGYGTGPGIGGQTA--AGQPG 259

Qy   51 VGTPAAAAAKAAAKAAQPGLVPGVG-----VAPGVGVAPGVGVAPGVGLAPGVG-----V 100
          |||   ||||||||| ||| |||| :    |     | |  |||||     | ||     |
Db  260 YGSQTGVGGQIGAGQPGYGSQPGIGGQTGAGQPGYGAQPGFGGQPGYGNQPGVGGQTGAG 319

Qy  101 APGVGVAPGVGVAPAIGPPEAQAAAAAAKAAKYGVGTPARAAAAKAAAKAAQPGLVPGVGVA 160
          |||||||| |||| |     :|:    |||| | ||| |||||  |||||| ||||| ||||
Db  320 QPGYGSQPGIGGQTGAGQP-GYGAQPGFGGQLGYGNQPGYGGQTGAGQPGYGSQPGYGGQ 378

Qy  161 PGVSVAPGVGVAPGLAPGVG-----VAPGVGVAPGVGVAPAIG 200
          || | || |||| |||| |||     ||||| | :|  |  :|:|
Db  379 TGAG-QPGYGVIPGFGGQPGIGGQQTAAGKPYGGQPGIGGSPYTG 422
```

RESULT 14
ID   O17434
AC   O17434;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   MINOR AMPULLATE SILK PROTEIN MiSP1 (FRAGMENT).
```

PRELIMINARY;   PRT;   988 AA.

```
OS   Nephila clavipes (Orb spider).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC   Araneomorphae; Entelegynae; Araneoidea; Tetragnathidae; Nephila.
OX   NCBI_TaxID=6915;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98200471; PubMed=9541398;
RA   Colgin M.A., Lewis R.V.;
RT   "Spider minor ampullate silk proteins contain new repetitive sequences
RT   and highly conserved non-silk-like 'spacer regions'.";
RL   Protein Sci. 7:667-672(1998).
DR   EMBL; AF027735; AAC14589.1; -.
DR   INTERPRO; IPR000817; -.
DR   PRINTS; PR00341; PRION.
FT   NON_TER          1          1
SQ   SEQUENCE    988 AA;  79082 MW;  461E03DF53F7085D CRC64;
```

```
Query Match            25.6%;  Score 260;  DB 5;  Length 988;
Best Local Similarity  36.8%;  Pred. No. 2.8e-09;
       Matches  75;  Conservative  5;  Mismatches  112;  Indels  12;  Gaps  4;

Qy       3 GFGVGVGGIPGVAGVPGVGVGGVPGVGISFENQAAAAAKAA------KYGVGTPAA    56
           |:|||:|  |  | |         ||   |   |   |       |  ||  |
Db     603 GYGAGAGAGAAAAATGAGGAGGYGRGAGAGAGAGAAGAGAGAGTGGAGYGGCGGSGYGAGAGAG  662

Qy      57 AAAKRAAKAAQPGLVPGVGVAPGVGVAPGVGL-APGVGVAPGVGVAPGVGVAPA  115
           |||| | |   :|| ||| ||  |:||    |  |:|| |   |:|| || | |
Db     663 AAAAAGAGGAGYGRGAGAGAGAAAGAGAAGAGAGGYGQGGGYGAGAGAGAAA    722

Qy     116 IGPPENQAAAAKAAKYG-VGTPAAAAAKAAKAAQCPGLVPGVGVAPGVGVA----PGVG  170
           |  | :| |   |  |   ||  |  :  |||      |  ||||  ||||
Db     723 AGAGGAGYSRGGRAGAAGAGAGAGAGAGCGAGGCYGGCGGYGAGAGAAGAAGAGAGSG  782

Qy     171 VAPGVGLAPGVGVAPGVGVAPGVG  194
           ||||| ||| ||  |||| |
Db     783 GAGGYGRGAGAAGAGAGAAGAG    806
```

```
RESULT 15
Q28100
ID   Q28100       PRELIMINARY;      PRT;    76 AA.
AC   Q28100;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85059254; PubMed=6150137;
RA   Rosenbloom J.;
RT   "Elastin: relation of protein and gene structure to disease.";
RL   Lab. Invest. 51:605-623(1984).
DR   EMBL; M31891; AAA96416.1; -.
DR   EMBL; M31893; AAA96416.1; JOINED.
DR   EMBL; M31892; AAA96416.1; JOINED.
FT   NON_TER          1          1
SQ   SEQUENCE    76 AA;  6619 MW;  B683379DAE87B202 CRC64;
```

```
Query Match            24.5%;  Score 249.5;  DB 6;  Length 76;
Best Local Similarity  60.8%;  Pred. No. 1.1e-09;
       Matches  62;  Conservative  2;  Mismatches  9;  Indels  29;  Gaps  4;

Qy      52 GTPAAA--AAKAAAKAAQFGLVPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPGVGVAPG  109
           | |||   || | ||| ||  ||| || |||| || | || ||:||| ||   |
Db       2 GVPAAAKSAAKAAAK-AQFRLGPGVGVAPGVGVVPGVGVVPGVGVVAPGNGLGPG------   54
```

```
Qy     110 VGVAPAIGPPEAQAAAAAKAAKAAKYGVGTPAAAAAKAAAKAAQF  151
           |     ||| ||||||||||||||||||     |||||||||||||
Db      55 -GV---------------IGVGIPAAAAAKAAAKAAQF   76
```

Search completed: April 24, 2001, 16:40:27
Job time: 417 sec

Tue Apr 24 16:55:21 2001

OM protein - protein search, using sw model

```
          Copyright    GenCore version 4.5
          Copyright (C) 1993 - 2000 Compugen Ltd.

Run on:   April 24, 2001, 16:38:15 ; Search time 115.25 Seconds
                                      (without alignments)
                                      1.984 Million cell updates/sec

Title:          US-09-340-736-3
Perfect score:  18
Sequence:       1 KAAK 4

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

ched:           390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters:    390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_0401:*
        1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
        2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
        3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
        4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
        5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
        6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
        7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
        8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
        9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
       10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
       11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
       12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
       13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
       14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
       15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
       16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
       17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
       18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
       19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
       20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
       21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
       22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	100.0	18	8	17	R91774	Proteinase A subst
2	18	100.0	18	8	21	B08137	Peptide modulating
3	18	100.0	18	8	21	B08138	Peptide modulating
4	18	100.0	18	8	21	B09144	Peptide modulating
5	18	100.0	18	9	16	R79675	Cyclin B/p33(cdc2)
6	18	100.0	18	9	21	Y57974	Xenopus histone H1
7	18	100.0	18	10	19	W48036	AE101 alanine scan
8	18	100.0	18	10	21	R04180	Monomer of repeat
9	18	100.0	18	11	21	B29025	SNP detection meth
10	18	100.0	18	11	21	Y84932	Peptide recognised
11	18	100.0	18	12	17	R88700	Allergen Alt a 12
12	18	100.0	12	18		W35172	Lys/Ala diastereom
13	18	100.0	12	18		W08390	Nucleic acid conde
14	18	100.0	12	18		W83888	Antipathogenic pep
15	18	100.0	12	19		W69198	Nucleic acid conde
16	18	100.0	12	19		W52869	Nucleic acid conde
17	18	100.0	12	21		B17423	Antipathogenic pep
18	18	100.0	13	21		I69144	Peptidomimetic of
19	18	100.0	13	9		P81754	Sequence of immuno
20	18	100.0	13	16		W21314	Glucagon precursor
21	18	100.0	13	16		R86018	Murine MHC class I
22	18	100.0	13	19		W33914	Interleukin-1 rece
23	18	100.0	13	20		Y39523	Pactolus protein a
24	18	100.0	14	17		W02565	AKAP79 AA38, 39 mu
25	18	100.0	14	17		W02564	AKAP79 (residues 3
26	18	100.0	14	21		B21902	Antimicrobial pro-
27	18	100.0	14	21		B14904	Human AKAP 79 pept
28	18	100.0	14	21		B14905	Mutant peptide AKA
29	18	100.0	15	15		R63701	BPI derived peptid
30	18	100.0	15	16		R75985	BPI protein segmen
31	18	100.0	15	16		R81030	BPI.23, BPI domain
32	18	100.0	15	16		R86479	BPI.23 for use in
33	18	100.0	15	16		R81110	Anti-fungal BPI pe
34	18	100.0	15	16		R75354	Bacterial permeabi
35	18	100.0	15	16		R76267	Bacterial permeabi
36	18	100.0	15	17		W07272	Amphiphilic antimi
37	18	100.0	15	17		W05876	Recombinant BPI pe
38	18	100.0	15	18		W29486	Peptide #7 derived
39	18	100.0	15	18		W29487	Peptide #8 derived
40	18	100.0	15	18		W29488	Peptide #9 derived
41	18	100.0	15	18		R43580	Bactericidal/perme
42	18	100.0	15	19		W66302	Magainin I or II a
43	18	100.0	15	19		W63327	Human BPI protein
44	18	100.0	15	20		Y34050	Histone HIS-3 deri
45	18	100.0	15	22		B53324	Peptide BPI 119.

ALIGNMENTS

```
RESULT   1
R91774
ID   R91774 standard; peptide; 8 AA.
XX
AC   R91774;
XX
DT   03-SEP-1996  (first entry)
XX
DE   Proteinase A substrate peptide #17.
XX
KW   Proteinase A; peptide substrate; activity; beer froth; colour;
KW   fluorescence; cleavage.
XX
OS   Synthetic.
XX
FH   Key            Location/Qualifiers
FH
FT   Modified-site  6
FT                  /label= OTHER
FT                  /note= "p-nitro-L-Phe"
XX
PN   JP08067694-A.
XX
PD   12-MAR-1995.
XX
PF   21-JUN-1995;  95JP-0154797.
XX
PR   21-JUN-1994;  94JP-0138704.
XX
PA   (SUNR ) SUNTORY LTD.
XX
DR   WPI; 1996-195571/20.
XX
PT   Peptide substrate for the determ. of proteinase A activity - useful
PT   to determine the stability of beer froth
```

```
XX
PS   Example 2; Page 7; 9pp; Japanese.
XX
CC   Peptides R91758-80 are specific examples of a novel generic proteinase A
CC   peptide substrate (R91757) and are modified with a substrate to develop a
CC   detectable colour or fluorescence after cleavage of the Phe-Phe bond by
CC   proteinase A. The peptides are used to determine proteinase A activity
CC   in a sample e.g. beer froth.
XX
SQ   Sequence  8 AA;

Query Match      100.0%;  Score 18;  DB 17;  Length 8;
Best Local Similarity  100.0%;  Pred. No. 3.2e+05;
Matches   4;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1 KAAK 4
       ||||
Db   1 kaak 4

RESULT 2
B08137
ID   B08137 standard; peptide; 8 AA.
XX
AC   B08137;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Peptide modulating activity of heparin, and other glycans.
XX
KW   Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW   cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW   cartilage differentiation; wound healing.
XX
OS   Synthetic.
XX
FH   Key         Location/Qualifiers
FT   Misc-difference 1..8
FT             /note= "this peptide may be repeated an
FT                    unspecified number of times"
XX
PN   WO200045831-A1.
XX
PD   10-AUG-2000.
XX
PF   02-FEB-2000; 2000WO-US02853.
XX
PR   02-FEB-1999;  99US-0118276.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
PI   San Antonio JD, Verrecchio A,  Schick BP;
XX
DR   WPI; 2000-543446/49.
XX
PT   Novel synthetic peptides with high affinity for glycoaminoglycans and
PT   proteoglycans, useful for modulating heparin, promoting cell
PT   attachment, modulating tumour metastasis and modulating wound healing -
XX
PS   Disclosure; Page 23; 76pp; English.
XX
CC   The present sequence represents a synthetic peptide which has a high
CC   affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC   in methods for modulating heparin or other glycoaminoglycans with
CC   anticoagulant activity, promoting cell attachment or adhesion to
CC   natural or synthetic surfaces (especially vein grafts), modulating
CC   tumour cell metastasis, modulating cartilage differentiation, targeting
CC   drugs to epithelial cell surfaces (or to other cells expressing
CC   proteoglycans), modulating enzymes that act on glycoaminoglycan
CC   substrates, affinity purification of bioactive sequences of a
CC   glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC   anti-coagulant functions mediated through glycoaminoglycans, and
CC   modulating wound healing. The peptide may also be used for blocking
CC   tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC   increase heparin half-life in circulation.
XX
SQ   Sequence  8 AA;

Query Match      100.0%;  Score 18;  DB 21;  Length 8;
Best Local Similarity  100.0%;  Pred. No. 3.2e+05;
Matches   4;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1 KAAK 4
       ||||
Db   4 kaak 7

RESULT 3
B08138
ID   B08138 standard; peptide; 8 AA.
XX
AC   B08138;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Peptide modulating activity of heparin, and other glycans.
XX
KW   Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW   cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW   cartilage differentiation; wound healing.
XX
OS   Synthetic.
XX
FH   Key         Location/Qualifiers
FT   Misc-difference 1..8
FT             /note= "this peptide may be repeated an
FT                    unspecified number of times"
XX
PN   WO200045831-A1.
XX
PD   10-AUG-2000.
XX
PF   02-FEB-2000; 2000WO-US02853.
XX
PR   02-FEB-1999;  99US-0118276.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
PI   San Antonio JD, Verrecchio A,  Schick BP;
XX
DR   WPI; 2000-543446/49.
XX
PT   Novel synthetic peptides with high affinity for glycoaminoglycans and
PT   proteoglycans, useful for modulating heparin, promoting cell
PT   attachment, modulating tumour metastasis and modulating wound healing -
XX
PS   Disclosure; Page 23; 76pp; English.
XX
CC   The present sequence represents a synthetic peptide which has a high
CC   affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC   in methods for modulating heparin or other glycoaminoglycans with
CC   anticoagulant activity, promoting cell attachment or adhesion to
CC   natural or synthetic surfaces (especially vein grafts), modulating
CC   tumour cell metastasis, modulating cartilage differentiation, targeting
CC   drugs to epithelial cell surfaces (or to other cells expressing
CC   proteoglycans), modulating enzymes that act on glycoaminoglycan
CC   substrates, affinity purification of bioactive sequences of a
CC   glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC   anti-coagulant functions mediated through glycoaminoglycans, and
CC   modulating wound healing. The peptide may also be used for blocking
CC   tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC   increase heparin half-life in circulation.
XX
SQ   Sequence  8 AA;
```

Query Match   100.0%; Score 18; DB 21; Length 8;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
              |||
Db   2 kaak 5

RESULT 4
B08144
ID   B08144 standard; peptide; 8 AA.
XX
AC   B08144;
XX
DT   04-DEC-2000 (first entry)
XX
DE   Peptide modulating activity of heparin, and other glycans.
XX
KW   Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW   cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW   cartilage differentiation; wound healing.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 1..8
FT   /note= "this peptide may be repeated an
FT   unspecified number of times"
XX
PN   WO200045831-A1.
XX
PD   10-AUG-2000.
XX
PF   02-FEB-2000; 2000WO-US02853.
XX
PR   02-FEB-1999; 99US-0118276.
XX
PA   (UVJE-) UNIV JEFFERSON THOMAS.
XX
PI   San Antonio JD, Verrecchio A, Schick BP;
XX
DR   WPI; 2000-543446/49.
XX
PT   Novel synthetic peptides with high affinity for glycoaminoglycans and
PT   proteoglycans, useful for modulating heparin, promoting cell
PT   attachment, modulating tumour metastasis and modulating wound healing -
XX
PS   Disclosure; Page 23; 76pp; English.
XX
CC   The present sequence represents a synthetic peptide which has a high
CC   affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC   in methods for modulating heparin or other glycoaminoglycans with
CC   anticoagulant activity, promoting cell attachment or adhesion to
CC   natural or synthetic surfaces (especially vein grafts), modulating
CC   tumour cell metastasis, modulating cartilage differentiation, targeting
CC   drugs to epithelial cell surfaces (or to other cells expressing
CC   proteoglycans), modulating enzymes that act on glycoaminoglycan
CC   substrates, affinity purification of bioactive sequences of a
CC   glycoaminoglycan, modifying endothelial cell pro-coaguant or
CC   anti-coagulant functions mediated through glycoaminoglycans, and
CC   modulating wound healing. The peptide may also be used for blocking
CC   tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC   increase heparin half-life in circulation.
XX
SQ   Sequence   8 AA;

Query Match   100.0%; Score 18; DB 21; Length 8;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
              ||||
Db   4 kaak 7

RESULT 5
R79675
ID   R79675 standard; peptide; 9 AA.
XX
AC   R79675;
XX
DT   26-FEB-1996 (first entry)
XX
DE   Cyclin B/p33(cdc2) phosphorylation site #2 in Xenopus histone H1.
XX
KW   Peptide library; phosphorylation site; protein kinase; substrate;
KW   inhibitor; competitor; cellular response; cell cycle control;
KW   immune response; transcriptional activation; cell development.
XX
OS   Synthetic.
XX
XX   WO9518823-A2.
PN
XX
PD   13-JUL-1995.
XX
PF   06-JAN-1995; 95WO-US00147.
XX
PR   07-JAN-1994; 94US-0178570.
XX
PA   (BETH-) BETH ISRAEL HOSPITAL ASSOC.
XX
PI   Cantley LC, Songyang Z;
XX
DR   WPI; 1995-255036/33.
XX
PT   Determ. of amino acid sequence of protein kinase phosphorylation
PT   site - by phosphorylation of peptide library and sequencing
PT   phospho:peptide(s) formed, also new substrates and their analogues
PT   for modulating or detecting protein kinase
XX
PS   Example 6; Page 34; 131pp; English.
XX
CC   An oriented degenerate peptide library of the amino acid formula
CC   R79661 was constructed to isolate the amino acid sequences at the
CC   phosphorylation sites of a protein kinase eg. protein kinase A,
CC   cyclin B/p33(cdc2), src family kinases, etc. Peptides which are
CC   phosphorylated are isolated and their amino acid sequences are compared
CC   to known substrate/inhibitor peptide sequences for that protein kinase.
CC   The peptides R79674-88 represent phosphorylation sites for the cell cycle
CC   control kinases cyclin B/p33(cdc2) or cyclin B/p33(CDK2). This peptide
CC   sequence is the second phosphorylation site in the Xenopus histone H1.
CC   The isolated peptides can be used to screen cpds. for effects on the
CC   protein kinase activity, generate antibodies to identify native kinase
CC   substrates, or modulate a variety of cellular responses in which protein
CC   kinases are involved eg. cell cycle control, immune response,
CC   transcriptional activation or cell development.
XX
SQ   Sequence   9 AA;

Query Match   100.0%; Score 18; DB 16; Length 9;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
              |||
Db   1 kaak 4

RESULT 6
Y57974
ID   Y57974 standard; peptide; 9 AA.

```
XX
AC   Y57974;
XX
DT   23-MAR-2000 (first entry)
XX
DE   Xenopus histone H1 peptide SEQ ID NO:28.
XX
KW   Protein kinase; phosphorylation site; signal transduction.
XX
OS   Xenopus sp.
XX
PN   US6004757-A.
XX
PD   21-DEC-1999.
XX
PF   06-JAN-1995;  95US-0369643.
XX
PR   07-JAN-1994;  94US-0178570.
XX
PA   (BETH-) BETH ISRAEL HOSPITAL ASSOC.
XX
PI   Cantley LC, Songyang Z;
XX
DR   WPI; 2000-096301/08.
XX
PT   Peptide substrate for a kinase, useful for determining substrate
PT   specificity -
XX
PS   Example 6; Column 31; 69pp; English.
XX
CC   The present invention describes a substrate for lck comprising a 9 amino
CC   acid peptide (I). Also described is a method of inhibiting kinase
CC   activity of lck by contacting it with (I) in vitro. The peptide is
CC   useful for determining substrate specificity of a protein kinase.
CC   Information on the substrate specificity of protein kinases in signal
CC   transduction would provide an insight into signal transduction
CC   mechanisms and could allow for the design of therapeutic agents. The
CC   present sequence represents a peptide used in the exemplification of
CC   the present invention.
XX
SQ   Sequence  9 AA;

Query Match       100.0%;  Score 18;  DB 21;  Length 9;
Best Local Similarity  100.0%;  Pred. No. 3.2e+05;
Matches   4;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1 KAAK 4
       ||||
Db   1 kaak 4

RESULT   7
W48036
ID   W48036 standard; peptide; 10 AA.
XX
AC   W48036;
XX
DT   12-JUN-1998 (first entry)
XX
DE   AE101 alanine scanning analogue effector compound SEQ ID NO:114.
XX
KW   Mammalian Ii key peptide; mammalian invariant chain protein; allergy;
KW   immune response; MHC class II; antigenic; autoimmune disease.
XX
OS   Synthetic.
XX
FH   Key        Location/Qualifiers
FT   Modified-site   1
FT              /note= "Acylated"
FT   Modified-site  10
FT              /note= "Amidated"
XX
```

```
PN   WO9749430-A1.
XX
PD   31-DEC-1997.
XX
PF   09-JUN-1997;  97WO-US09993.
XX
PR   26-JUN-1996;  96US-0670605.
XX
PA   (ANTI-) ANTIGEN EXPRESS INC.
XX
PI   Adams S,  Humphreys RE,  Xu M;
XX
DR   WPI; 1998-076917/07.
XX
PT   New mammalian invariant chain protein (Ii) key peptide(s) - used for
PT   modulation of immune response, e.g. for treating malignant, allergic
PT   or autoimmune disease or allograft rejection
XX
PS   Example 2; Page 40; 149pp; English.
XX
CC   The present sequence represents an AE101 analogue effector compound
CC   used in the present invention. The present invention describes
CC   a mammalian invariant chain protein (Ii) key peptide of sequence
CC   LRMKLPKPPKPVSKMR and modifications with the exclusion of peptide
CC   YRMKLPKPPKPVSKMR. MHC class II molecules are synthesised in the
CC   endoplasmic reticulum with their antigenic peptide sites blocked by the
CC   invariant chain protein (Ii). The products and method can be used for
CC   the modulation of an immune response for therapeutic or diagnostic
CC   purposes. The enhancement of immunity can be used in the treatment of
CC   e.g. malignant or allergic disease. The immunosuppression can be used
CC   for the treatment of autoimmune disease, e.g. rheumatoid arthritis,
CC   multiple sclerosis, diabetes mellitus, lupus erythematosus, and
CC   psoriasis or allograft rejection.
XX
SQ   Sequence  10 AA;

Query Match       100.0%;  Score 18;  DB 19;  Length 10;
Best Local Similarity  100.0%;  Pred. No. 1e+02;
Matches   4;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1 KAAK 4
       ||||
Db   7 kaak 10

RESULT   8
R04180
ID   R04180 standard; protein; 11 AA.
XX
AC   R04180;
XX
DT   10-SEP-1990 (first entry)
XX
DE   Monomer of repetative primary structure of a synthetic fibrous
DE   protein.
XX
KW   Collagen; elastin; keratin; troponin C; silk; dopa protein;
KW   synthetic skin; cosmetics; bioadhesive.
XX
OS   Synthetic.
XX
FH   Key        Location/Qualifiers
FT   Misc-difference   9
FT              /label=May be Phe or Tyr.
XX
PN   WO9003438-A.
XX
PD   05-APR-1990.
XX
PF   06-SEP-1989;  89WO-0003839.
XX
PR   30-SEP-1988;  88US-0251714.
XX
```

```
XX
PA  (ALSC ) ALLIED SIGNAL INC.
XX
PI  Goldberg I, Salerno AJ;
XX
DR  WPI; 1990-132274/17.
XX
PT  New bacterial strains for heterologous gene expression -
PT  contg. elements for initiating activity, retarding proteolysis
PT  and stabilising heterologous genes.
XX
PS  Disclosure; ; 1pp; English.
XX
CC  The polymer of the peptide may act as an analogue to peptides
CC  contg. a repetitive or quasi-repetitive structure eg. collagen,
CC  elastin, keratin, troponin C, dopa proteins, silk proteins,
CC  bioadhesive proteins and insect cuticle proteins. These products
CC  may be useful in the manufacture of fibrous products, synthetic
CC  skin and cosmetic additives.
XX
SQ  Sequence  11 AA;

Query Match       100.0%;   Score 18;  DB 11;  Length 11;
Best Local Similarity  100.0%;  Pred. No. 1.1e+02;
Matches   4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   1 KAAK 4
       ||||
Db   5 kaak 8

RESULT   9
ID  B29025 standard; peptide; 11 AA.
XX
AC  B29025;
XX
DT  26-JAN-2001 (first entry)
XX
DE  SNP detection method labelling antibody peptide SEQ ID NO: 1.
XX
KW  Single nucleotide polymorphism; SNP; antibody; label; rabbit; cancer;
KW  cardiovascular disease; neurological disease; metabolic disease;
KW  autoimmune disease.
XX
OS  Oryctolagus cuniculus.
XX
PN  WO200056926-A2.
XX
PD  28-SEP-2000.
XX
PF  16-MAR-2000; 2000WO-US05950.
XX
PR  19-MAR-1999; 99US-0272970.
XX
PA  (VALI-) VALIGENE CORP.
PA  (TREI/) TREICH I.
XX
PI  Treich I, Iris FJM, Pourny J;
XX
DR  WPI; 2000-647157/62.
XX
PS  Disclosure; Page 33; 62pp; English.
XX
CC  The present invention relates to methods for detecting single nucleotide
CC  polymorphisms (SNPs) using peptide-labelled oligonucleotides and antibody
CC  arrays. Oligonucleotide probes with identifiable labels are used in the
CC  ValiGeneSM Mutation Screening Peptide-Linked (VGMS-PL) method. These
```

```
CC  labels may be peptides or antibodies. The present sequence is an example
CC  of the peptide which may be used as a label. The method is useful in
CC  detecting SNPs which may be related to numerous inherited human diseases,
CC  including cancers, neurological diseases, metabolic diseases,
CC  cardiovascular diseases and autoimmune diseases. In addition, it can be
CC  used to detect and diagnose Lyme disease, tuberculosis and sexually
CC  transmitted diseases.
XX
SQ  Sequence  11 AA;

Query Match       100.0%;   Score 18;  DB 21;  Length 11;
Best Local Similarity  100.0%;  Pred. No. 1.1e+02;
Matches   4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   1 KAAK 4
       ||||
Db   7 kaak 10

RESULT  10
ID  Y84932 standard; peptide; 11 AA.
XX
AC  Y84932;
XX
DT  21-AUG-2000 (first entry)
XX
DE  Peptide recognised by rabbit IgG antibodies.
XX
KW  IgG antibody; cDNA library; nucleic acid sorting; gene expression.
XX
OS  Oryctolagus cuniculus.
XX
PN  WO200023622-A1.
XX
PD  27-APR-2000.
XX
PF  15-OCT-1999;  99WO-US23906.
XX
PR  16-OCT-1998;  98US-0174328.
XX
PA  (VALI-) VALIGENE CORP.
XX
PI  Iris FJ, Pourny J;
XX
DR  WPI; 2000-339716/29.
XX
PT  Peptide-Labeled oligonucleotide methods for manipulating, e.g.
PT  labeling, isolating or screening, populations of nucleic acids involve
PT  the use of distinguishable and identifiable peptide tags linked to
PT  oligonucleotide primers -
XX
PS  Disclosure; Page 40; 61pp; English.
XX
CC  The present sequence represents a peptide which is recognised by rabbit
CC  IgG antibodies. The antibody that recognises the present peptide may
CC  be used in the method of the invention. The specification describes a
CC  method for sorting a mixture of nucleic acids derived from cDNA
CC  libraries. The method comprises labeling DNA from each of the libraries
CC  by polymerase chain reaction (PCR) using distinguishably-labelled
CC  oligonucleotide primers for each library, contacting differently
CC  labelled DNA so hybridisation can occur, and sorting DNA using one or
CC  more molecules, each capable of binding one of the labels. The method
CC  is useful for labelling, sorting, comparing, detecting and screening
CC  multiple, complex populations of nucleic acids. The method is further
CC  useful for monitoring gene expression. The method is also useful for
CC  increasing and supplementing the analytical powers of other techniques
CC  of manipulating complex cDNA populations.
XX
SQ  Sequence  11 AA;
```

```
Query Match          100.0%;  Score 18;  DB 21;  Length 11;
Best Local Similarity 100.0%;  Pred. No. 1.1e+02;
Matches    4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 KAAK 4
        ||||
Db    7 kaak 10

RESULT 11
R88700
ID   R88700 standard; Peptide; 12 AA.
XX
AC   R88700;
XX
DT   19-AUG-1996  (first entry)
XX
DE   Allergen Alt a 12 T-cell epitope 1.
XX
KW   Alt a 45; Alt a 12; allergen; epitope; immunoglobulin E; IgE;
KW   detection; immunotolerance; anergy.
XX
OS   Alternaria alternata.
XX
PN   AT940203B-A.
XX
PD   15-NOV-1995.
XX
PF   02-NOV-1994;  94AT-0002038.
XX
PR   02-NOV-1994;  94AT-0002038.
XX
PA   (BIOM-) BIOMAY PRODN & HANDELSGES MBH.
XX
PI   Achatz G, Breitenbach M, Ebner C, Kraft D, Lechenauer E;
PI   Oberkofler H, Frillinger B, Simon B, Unger A;
XX
DR   WPI; 1996-040555/05.
XX
PT   Recombinant DNA encoding allergens of Alternaria alternata - useful
PT   in diagnosis and treatment of A. alternata allergies
XX
PS   Claim 1; Page 15; 21pp; German.
XX
CC   R88700-01 are T-cell epitopes derived from the Alt a 12 allergen protein
CC   (R88692) isolated from Alternaria alternata. Peptide epitopes from Alt a
CC   12 and Alt a 45 (R88672) are useful as diagnostic reagents; e.g. for in
CC   vitro detection of allergy caused by Alt a 45 and 12 (by reaction with
CC   IgE in serum). They can also detect cellular reaction to the specified
CC   allergens (from their stimulatory or inhibitory effect on this reaction).
CC   The peptides can also be used therapeutically to induce immunotolerance
CC   or anergy of T lymphocytes.
XX
SQ   Sequence   12 AA;

Query Match          100.0%;  Score 18;  DB 17;  Length 12;
Best Local Similarity 100.0%;  Pred. No. 1.2e+02;
Matches    4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 KAAK 4
        ||||
Db    7 kaak 10

RESULT 12
W35172
ID   W35172 standard; peptide; 12 AA.
XX
AC   W35172;
XX
DT   14-APR-1998  (first entry)
XX
```

```
DE   Lys/Ala diastereomer peptide [D]-A3,4,8,10-K4A8.
XX
KW   Lys/Ala diastereomer peptide; infection; therapy; excitatory neurotoxin;
KW   Honey bee venom; pardaxin; cytolytic activity; cancer;
KW   non-haemolytic; preservative; agricultural produce; bacterial cell lysis;
KW   agricultural pesticide; cell wall lysis.
XX
OS   Synthetic.
XX
FH   Key              Location/Qualifiers
FT   Misc-difference  3
FT                    /note= "D-form residue"
FT   Misc-difference  4
FT                    /note= "D-form residue"
FT   Misc-difference  8
FT                    /note= "D-form residue"
FT   Misc-difference  10
FT                    /note= "D-form residue"
FT   Modified-site    12
FT                    /note= "C-terminal amide"
XX
PN   WO9731019-A2.
XX
PD   28-AUG-1997.
XX
PF   20-FEB-1997;  97WO-IL00066.
XX
PR   22-FEB-1996;  96IL-0117223.
XX
PA   (YEDA ) YEDA RES & DEV CO LTD.
XX
PI   Oren Z, Shai Y;
XX
DR   WPI; 1997-435088/40.
XX
PT   Peptide(s) having selective cytolytic activity - against pathogens
PT   and malignant cells, but no haemolytic activity, used for treating
PT   infections and cancer
XX
PS   Example 4; Page 45; 80pp; English.
XX
CC   This sequence represents a Lys/Ala diastereomer peptide of the
CC   invention. The peptides of the invention have: (a) cytolytic activity on
CC   pathogenic cells (pathogens and malignant cells not naturally present in
CC   the body); but (b) no haemolytic activity, or such activity only at a
CC   concentration significantly higher than that at which they lyse
CC   pathogens. The peptides, their complexes and mixtures are used to treat
CC   infections (caused by bacteria, fungi, protozoa, mycoplasma or viruses)
CC   or cancer, in human and veterinary medicine. Also, they can be used as
CC   preservatives for food, cosmetics and agricultural produce, or as
CC   agricultural pesticides. The absence of haemolytic activity (associated
CC   with disturbance of alpha-helical structures) means that the peptides
CC   have few if any toxic effects, and those that include D-aa will have
CC   increased resistance to proteolytic degradation. Non-haemolytic,
CC   cytotoxic random copolymers of pardaxin, each has a specific spectrum of
CC   activity, allowing selection of agents for particular applications. Since
CC   these random copolymers induce total lysis of bacterial cell walls,
CC   resistance to them is unlikely to develop.
XX
SQ   Sequence   12 AA;

Query Match          100.0%;  Score 18;  DB 18;  Length 12;
Best Local Similarity 100.0%;  Pred. No. 1.2e+02;
Matches    4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 KAAK 4
        ||||
Db    9 kaak 12

RESULT 13
W08390
```

ID   WO8390 standard; peptide; 12 AA.
XX
AC   WO8390;
XX
DT   19-SEP-1997 (first entry)
XX
DE   Nucleic acid condensing peptide consensus sequence III.
XX
KW   Synthetic virus like particle; nucleic acid; condensing peptide;
KW   heteropeptide; polydispersion; transfection; mammalian cell;
KW   gene therapy; self assembly; consensus.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 1
FT   /label= Lys, Arg
FT   Misc-difference 2
FT   /label= Ala, Thr
FT   Misc-difference 10
FT   /label= Pro, Ala, Ser
FT   Misc-difference 12
FT   /label= Lys, Thr, Val
XX
PN   WO9641606-A2.
XX
PD   27-DEC-1996.
XX
PF   10-JUN-1996;   96WO-GB01396.
XX
PR   07-JUN-1996;   96US-0560231.
PR   08-JUN-1995;   95US-0000124.
PR   30-JUN-1995;   95GB-0013399.
PR   21-SEP-1995;   95GB-0019304.
PR   25-SEP-1995;   95US-0004285.
PR   19-DEC-1995;   95GB-0025955.
PR   19-DEC-1995;   95US-0008952.
PR   12-FEB-1996;   96US-0011531.
XX
PA   (THER-) THEREXSYS LTD.
XX
PI   Craig RK, Cunliffe VT, Thatcher DR, Welsh JH, Wilks PE;
XX
DR   WPI; 1997-065278/06.
XX
PS   Claim 36; Page 123; 193pp; English.
XX
CC   A novel synthetic virus like particle (SVLP), comprises a
CC   recombinant nucleic acid molecule and a non-covalently associated
CC   set of nucleic acid condensing peptides (NACP), e.g. a peptide
CC   comprising the present consensus sequence, where each NACP is a
CC   heteropeptide, and the NACP set has low polydispersion. The SVLP,
CC   which is for transfecting a nucleic acid molecule into a mammalian
CC   cell, can be used for gene therapy. It is self assembling, and can
CC   be designed to target a particular cell or tissue type and deliver
CC   the nucleic acid molecule into its chromosomal or extrachromosomal
CC   sequences.
XX
SQ   Sequence 12 AA;

Query Match   100.0%; Score 18; DB 18; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
        ||||
Db   6 KAAK 9

RESULT 14
WB2888
ID   WB2888 standard; peptide; 12 AA.
XX
AC   WB2888;
XX
DT   19-MAY-1999 (first entry)
XX
DE   Antipathogenic peptide.
XX
KW   Non-haemolytic; cytolytic; selective cytolytic activity; pathogen;
KW   cancer; infection; disinfectant; contact lens wetting solution;
KW   preservative; pesticide; fungicide; bactericide.
XX
OS   Synthetic.
XX
PN   WO9837090-A1.
XX
PD   27-AUG-1998.
XX
PF   19-FEB-1998;   98WO-IL00081.
XX
PR   20-FEB-1997;   97WO-IL00066.
XX
PA   (YEDA) YEDA RES & DEV CO LTD.
XX
PI   Oren Z, Shai Y;
XX
DR   WPI; 1998-594464/50.
XX
PT   New non-haemolytic cytolytic agent useful in treating cancer or
PT   infections - is a peptide comprising a moiety which disrupts the
PT   continuity of an alpha-helical structure
XX
PS   Example 4; Page 44; 126pp; English.
XX
CC   The present peptide is used to produce the agents of the invention. The
CC   specification describes a non-haemolytic, cytolytic agent, which is a
CC   peptide, a complex of bundled peptides, a mixture of peptides or a random
CC   peptide copolymer. The agent has a selective cytolytic activity on
CC   pathogenic cells. The agent is selected from a cyclic derivative of a
CC   peptide which has a net positive charge greater than 1, comprises L-amino
CC   acid residues and/or D-amino acid residues and comprises an alpha-helix
CC   breaker moiety, or a peptide (or cyclic derivative of this) which has a net
CC   (comprises L-amino acid residues and D-amino acid residues, has a net
CC   positive charge greater than 1 and has an amino acid sequence such that
CC   a corresponding amino acid sequence comprising only L-amino acid residues
CC   is not found in nature. The cytolytic agents may be used for treatment of
CC   cancer or for treatment of several diseases caused by pathogens,
CC   including bacterial, fungal, viral, mycoplasma and protozoan infections.
CC   They may be used in both human and veterinary medicine. They may also be
CC   used as disinfectants for destruction of microorganisms, i.e. in
CC   solutions for wetting contact lenses, as preservatives, e.g., in the
CC   cosmetic and food industries, as pesticides (e.g. fungicides or
CC   bactericides) or for preservation of agricultural products.
XX
SQ   Sequence 12 AA;

Query Match   100.0%; Score 18; DB 19; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
        ||||
Db   9 KAAK 12

RESULT 15
W69198
ID   W69198 standard; peptide; 12 AA.
XX
AC   W69198;

```
XX
DT   09-OCT-1998 (first entry)
XX
DE   Nucleic acid condensing peptide #3.
XX
KW   Nucleic acid condensing peptide; cell transfection; gene delivery system.
XX
OS   Synthetic.
XX
FH   Key            Location/Qualifiers
FT   Misc-difference 1
FT                  /label= Lys, Arg
FT   Misc-difference 2
FT                  /label= Ala, Val, Thr
FT   Misc-difference 10
FT                  /label= Pro, Ala, Ser
FT   Misc-difference 12
FT                  /label= Ala, Lys, Thr, Val
XX
PN   WO9828625-A1.
XX
PD   02-JUL-1998.
XX
PF   23-DEC-1997;   97WO-GB03523.
XX
PR   27-DEC-1996;   96GB-0026992.
PR   23-DEC-1996;   96US-0033908.
XX
PA   (COBR-) COBRA THERAPEUTICS LTD.
XX
PI   Thatcher DR,  Wilks PE;
XX
DR   WPI; 1998-377790/32.
XX
PT   Transfection of cells with nucleic acid(s) - comprises use of
PT   peptide(s) screened for interaction with nucleic acids on sensor
PT   chip by surface plasmon resonance
XX
PS   Disclosure; Page 23; 75pp; English.
XX
CC   This sequence represents a nucleic acid condensing peptide that was
CC   identified using the method of the invention. The method is for screening
CC   test peptides for the ability to optimally transfect cells with a nucleic
CC   acid molecule (NAM), and comprises: (a) detecting a change in the surface
CC   plasmon resonance (SPR) of a nucleic acid immobilised on a sensor chip
CC   and exposed to a solution of a test peptide, where the change in SPR
CC   occurs upon binding of the peptide to and dissociation of the peptide
CC   from the immobilised nucleic acid, to permit calculation of the
CC   equilibrium constant (K(d)) (or apparent dissociation constant (kd)); and
CC   (b) selecting the peptide having K(d), with a value of 1x10-12 to 1x10-6
CC   or kd, with a value of 1x10-6 to 1x10-1.The methods can be used for
CC   optimising the efficiency of transfecting cells with NAMs. They permit
CC   the selection of peptides having binding capacity for nucleic acid to
CC   provide a gene delivery complex that is stable in the bloodstream,
CC   targetable to selected tissue types, and capable of efficient transport
CC   into the cytoplasm and to the nucleus.
XX
SQ   Sequence  12 AA;

Query Match       100.0%;  Score 18;  DB 19;  Length 12;
Best Local Similarity  100.0%; Pred. No. 1.2e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 KAAK 4
           ||||
Db     6 kaak 9

Search completed: April 24, 2001, 16:38:17
Job time: 417 sec
```

```
                GenCore version 4.5
        Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:      April 24, 2001, 16:36:22 ; Search time 62.39 Seconds
                                       (without alignments)
                                       1.232 Million cell updates/sec

Title:          US-09-340-736-3
Perfect score:  18
Sequence:       1 KAAK 4

Scoring table:  BLOSUM62
Gapop 10.0 , Gapext 0.5

tched:          185757 segs, 19210857 residues

Total number of hits satisfying chosen parameters:   185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Issued_Patents_AA:*
          1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
          2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
          3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
          4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
          5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
          6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                          SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	18	100.0	4	1	US-07-694-983-8	Sequence 8, Appli
2	18	100.0	4	1	US-08-951-364-3	Sequence 3, Appli
3	18	100.0	9	3	US-08-178-570-28	Sequence 28, Appl
4	18	100.0	9	3	US-08-369-643-28	Sequence 28, Appl
5	18	100.0	9	5	PCT-US95-00147-28	Sequence 28, Appl
6	18	100.0	10	5	US-08-968-676-114	Sequence 114, App
7	18	100.0	13	5	PCT-US95-04121-38	Sequence 38, Appl
8	18	100.0	14	2	US-08-503-2208-37	Sequence 37, Appl
9	18	100.0	14	2	US-08-503-2268-38	Sequence 38, Appl
10	18	100.0	14	2	US-08-860-150-18	Sequence 18, Appl
11	18	100.0	14	3	US-09-338-132-18	Sequence 18, Appl
12	18	100.0	14	3	US-08-721-458B-37	Sequence 37, Appl
13	18	100.0	14	3	US-08-721-458B-38	Sequence 38, Appl
14	18	100.0	15	1	US-08-311-611A-24	Sequence 24, Appl
15	18	100.0	15	1	US-08-311-611A-115	Sequence 115, App
16	18	100.0	15	1	US-08-372-783-24	Sequence 24, Appl
17	18	100.0	15	1	US-08-372-783-115	Sequence 115, App
18	18	100.0	15	1	US-08-372-105-24	Sequence 24, Appl
19	18	100.3	15	1	US-08-372-105-115	Sequence 115, App
20	18	100.0	15	1	US-08-306-473A-24	Sequence 24, Appl
21	18	100.0	15	1	US-08-306-473A-115	Sequence 115, App
22	18	100.0	15	1	US-08-209-762-24	Sequence 24, Appl
23	18	100.0	15	1	US-08-473-344-24	Sequence 24, Appl
24	18	100.0	15	2	US-08-621-803-24	Sequence 24, Appl
25	18	100.0	15	2	US-08-485-445A-24	Sequence 24, Appl
26	18	100.0	15	2	US-08-485-445A-115	Sequence 115, App
27	18	100.0	15	2	US-08-338-882-7	Sequence 7, Appli

28	18	100.0	15	3	US-09-041-889-22	Sequence 22, Appl
29	18	100.0	15	3	US-09-041-889-23	Sequence 23, Appl
30	18	100.0	15	3	US-09-041-889-24	Sequence 24, Appl
31	18	100.0	15	3	US-09-119-263-24	Sequence 24, Appl
32	18	100.0	15	3	US-09-119-263-115	Sequence 115, App
33	18	100.0	15	3	US-08-837-058-22	Sequence 22, Appl
34	18	100.0	15	3	US-08-837-058-23	Sequence 23, Appl
35	18	100.0	15	4	US-08-837-058-24	Sequence 24, Appl
36	18	100.0	15	4	US-08-657-162-24	Sequence 24, Appl
37	18	100.0	15	4	US-08-657-162-115	Sequence 115, App
38	18	100.0	15	4	US-09-224-480-24	Sequence 24, Appl
39	18	100.0	15	5	US-09-224-480-115	Sequence 115, App
40	18	100.0	15	5	PCT-US94-02465-24	Sequence 24, Appl
41	18	100.0	15	5	PCT-US95-00498-24	Sequence 24, Appl
42	18	100.0	15	5	PCT-US95-00498-115	Sequence 115, App
43	18	100.0	15	5	PCT-US95-00656-24	Sequence 24, Appl
44	18	100.0	15	5	PCT-US95-00656-115	Sequence 115, App
45	18	100.0	19	4	US-09-166-930A-5	Sequence 5, Appli

```
                          ALIGNMENTS

RESULT   1
US-07-694-983-8
; Sequence 8, Application US/07694983
; Patent No. 5432260
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Philip D.
;  TITLE OF INVENTION: HIGH AFFINITY MANNOSE RECEPTOR
;  TITLE OF INVENTION: LIGANDS
;  NUMBER OF SEQUENCES: 19
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Irell & Manella
;   STREET: 545 Middlefield Road, Suite 200
;   CITY: Menlo Park
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94025
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/694,983
;   FILING DATE: 19910503
;   CLASSIFICATION: 530
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Murashige, Kate H.
;   REGISTRATION NUMBER: 29,959
;   REFERENCE/DOCKET NUMBER: 9500-0039.00
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-327-7250
;   TELEFAX: 415-327-2951
;   TELEX: 706141
;  INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH: 4 amino acids
;    TYPE: AMINO ACID
;    STRANDEDNESS: single
;    TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-07-694-983-8

Query Match             100.0%;  Score 18;  DB 1;  Length 4;
Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 KAAK 4
             ||||
Db         1 KAAK 4
```

RESULT 2
US-08-911-364-3
; Sequence 3, Application US/08911364
; Patent No. 5969106
; GENERAL INFORMATION:
; APPLICANT: ROTHSTEIN, Aser
; APPLICANT: KEELY, Fred W.
; APPLICANT: ROTHSTEIN, Steven J.
; TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
; TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FOLEY & LARDNER
;   STREET: 3000 K Street, N.W.
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20007-5109
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/911,364
;   FILING DATE: 07-AUG-1997
;   CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/023,552
;   FILING DATE: 07-AUG-1996
; ATTORNEY/AGENT INFORMATION:
;   NAME: Bent, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 041082/0104
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 672-5300
;   TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 4 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-911-364-3

Query Match   100.0%; Score 18; DB 2; Length 4;
est Local Similarity 100.0%; Pred No. 1.4e+05;
tches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
         ||||
Db   1 KAAK 4

RESULT 3
US-08-178-570-28
; Sequence 28, Application US/08178570
; Patent No. 5532167
; GENERAL INFORMATION:
; APPLICANT: Lewis C. Cantley
; APPLICANT: Zhou Song yang
; TITLE OF INVENTION: Substrate Specificity of Protein Kinases
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: LAHIVE & COCKFIELD
;   STREET: 60 STATE STREET, suite 510
;   CITY: BOSTON
;   STATE: MASSACHUSETTS
;   COUNTRY: USA

;   ZIP: 02109-1875
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/178,570
;   FILING DATE: JANUARY 7, 1994
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: DeConti, Giulio A., Jr.
;   REGISTRATION NUMBER: 31,503
;   REFERENCE/DOCKET NUMBER: BBI-004
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 227-7400
;   TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 9 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   FRAGMENT TYPE: internal
US-08-178-570-28

Query Match   100.0%; Score 18; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
         ||||
Db   1 KAAK 4

RESULT 4
US-08-369-643-28
; Sequence 28, Application US/08369643A
; Patent No. 6004757
; GENERAL INFORMATION:
; APPLICANT: Cantley, Lewis C.
; APPLICANT: Songyang, Zhou
; TITLE OF INVENTION: Substrate Specificity of Protein Kinases
; FILE REFERENCE: CNS-001CP
; CURRENT APPLICATION NUMBER: US/08/369,643A
; CURRENT FILING DATE: 1995-01-06
; EARLIER APPLICATION NUMBER: US 08/178,570
; EARLIER FILING DATE: 1994-01-07
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
;   LENGTH: 9
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
; FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Xenopus histone
;   OTHER INFORMATION: H1
US-08-369-643-28

Query Match   100.0%; Score 18; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
         ||||
Db   1 KAAK 4

RESULT 5
PCT-US95-00147-28
; Sequence 28, Application PC/TUS9500147

GENERAL INFORMATION:
; APPLICANT:
;   TITLE OF INVENTION: Substrate Specificity of Protein Kinases
;   NUMBER OF SEQUENCES: 88
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: LAHIVE & COCKFIELD
;   STREET: 60 STATE STREET, suite 510
;   CITY: BOSTON
;   STATE: MASSACHUSETTS
;   COUNTRY: USA
;   ZIP: 02109-1875
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: ASCII text
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/00147
;   FILING DATE:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/178,570
;   FILING DATE: JANUARY 7, 1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: DeConti, Giulio A., Jr.
;   REGISTRATION NUMBER: 31,503
;   REFERENCE/DOCKET NUMBER: BBI-004CPPC
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 227-7400
;   TELEFAX: (617) 227-5941
;   INFORMATION FOR SEQ ID NO: 28:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 9 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   FRAGMENT TYPE: internal
PCT-US95-00147-28

Query Match   100.0%; Score 18; DB 5; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
          ||||
Db   1 KAAK 4

RESULT 6
US-08-968-676-114
; Sequence 114, Application US/08968676
; Patent No. 5919639
; GENERAL INFORMATION:
;   APPLICANT: Humphreys, Robert E
;   APPLICANT: Adams, Sharlene
;   APPLICANT: Xu, Minzhen
;   TITLE OF INVENTION: IMMUNOTHERAPY BY MODULATION OF ANTIGEN
;   NUMBER OF SEQUENCES: 165
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Kevin M. Farrell, P.C.
;   STREET: P.O. Box 999
;   CITY: York Harbor
;   STATE: ME
;   COUNTRY: USA
;   ZIP: 03911
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/968,676
;   FILING DATE:

;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Farrell, Kevin M
;   REGISTRATION NUMBER: 35,505
;   REFERENCE/DOCKET NUMBER: REH-9601
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (207) 363-0558
;   TELEFAX: (207) 363-0528
; INFORMATION FOR SEQ ID NO: 114:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 10 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-968-676-114

Query Match   100.0%; Score 18; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 74;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
          ||||
Db   7 KAAK 10

RESULT 7
PCT-US95-04121-38
; Sequence 38, Application PCT/US9504121
; GENERAL INFORMATION:
;   APPLICANT:
;   TITLE OF INVENTION: Haptenated Peptides and Uses Thereof
;   NUMBER OF SEQUENCES: 62
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/04121
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/222,206
;   FILING DATE: April 1, 1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Vanstone, Darlene A.
;   REGISTRATION NUMBER: 35,279
;   REFERENCE/DOCKET NUMBER: 079.2PCT
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 466-6000
;   TELEFAX: (617) 466-6010
; INFORMATION FOR SEQ ID NO: 38:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 13 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   FRAGMENT TYPE: internal
PCT-US95-04121-38

Query Match   100.0%; Score 18; DB 5; Length 13;
Best Local Similarity 100.0%; Pred. No. 93;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
          ||||
Db   4 KAAK 7

RESULT 8
US-08-503-226B-37
; Sequence 37, Application US/08503226B
; Patent No. 5871945
; GENERAL INFORMATION:
; APPLICANT: Lockerbie, Robert Owen, et al.
; TITLE OF INVENTION: Calcineurin Inhibitory Compounds and
; TITLE OF INVENTION: Anchoring Protein
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/503,226B
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/404,731
; FILING DATE: 15-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,227
; FILING DATE: 23-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32861
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-6448
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-503-226B-37

Query Match   100.0%; Score 18; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAK 4
       ||||
Db 6 KAAK 9

RESULT 9
US-08-503-226B-38
; Sequence 38, Application US/08503226B
; Patent No. 5871945
; GENERAL INFORMATION:
; APPLICANT: Lockerbie, Robert Owen, et al.
; TITLE OF INVENTION: Calcineurin Inhibitory Compounds and
; TITLE OF INVENTION: Anchoring Protein
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606

RESULT 10
US-08-860-150-18
; Sequence 18, Application US/08860150B
; Patent No. 5981205
; GENERAL INFORMATION:
; APPLICANT: Hemmings, Brian A.
; APPLICANT: Millward, Thomas A.
; TITLE OF INVENTION: Nuclear DBF2-Related (NDR) Kinases
; FILE REFERENCE: 4-20265/A/PCT
; CURRENT APPLICATION NUMBER: US/08/860,150B
; CURRENT FILING DATE: 1997-06-19
; EARLIER APPLICATION NUMBER: PCT/ZEP95/05052
; EARLIER FILING DATE: 1995-12-20
; EARLIER APPLICATION NUMBER: 94810746.1
; EARLIER FILING DATE: 1994-12-22
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(14)
; OTHER INFORMATION: Predicted carboxy terminus of human Ndr
US-08-860-150-18

Query Match   100.0%; Score 18; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 KAAK 4
       ||||
Db   11 KAAK 14

RESULT 11
US-09-338-132-18
; Sequence 18, Application US/09338132
; Patent No. 6040164
; GENERAL INFORMATION:
; APPLICANT: Hemmings, Brian A.
; APPLICANT: Millward, Thomas A.
; TITLE OF INVENTION: Nuclear DBF2-Related (NDR) Kinases
; FILE REFERENCE: 4-20265/A/PCT
; CURRENT APPLICATION NUMBER: US/09/338,132
; CURRENT FILING DATE: 1999-06-22
; EARLIER APPLICATION NUMBER: 08/860,150
; EARLIER FILING DATE: 1997-06-19
; EARLIER APPLICATION NUMBER: PCT/EP95/05052
; EARLIER FILING DATE: 1995-12-20
; EARLIER APPLICATION NUMBER: 94810746.1
; EARLIER FILING DATE: 1994-12-22
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(14)
; OTHER INFORMATION: Predicted carboxy terminus of human Ndr
US-09-338-132-18

Query Match   100.0%; Score 18; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 KAAK 4
       ||||
Db   11 KAAK 14

RESULT 12
US-08-721-458B-37
; Sequence 37, Application US/08721458B
; Patent No. 6107104
; GENERAL INFORMATION:
; APPLICANT: Lockerbie, Robert Owen, et al.
; TITLE OF INVENTION: Calcineurin Inhibitory Compounds
; TITLE OF INVENTION: and Anchoring Protein
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,458B
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/404,731
; FILING DATE: 15-MAR-1995
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/344,227
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/503,226
; FILING DATE: 17-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33276
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-6448
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-721-458B-37

Query Match   100.0%; Score 18; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 KAAK 4
       ||||
Db   6 KAAK 9

RESULT 13
US-08-721-458B-38
; Sequence 38, Application US/08721458B
; Patent No. 6107104
; GENERAL INFORMATION:
; APPLICANT: Lockerbie, Robert Owen, et al.
; TITLE OF INVENTION: Calcineurin Inhibitory Compounds
; TITLE OF INVENTION: and Anchoring Protein
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,458B
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/404,731
; FILING DATE: 15-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,227
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/503,226
; FILING DATE: 17-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33276
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-6448
; INFORMATION FOR SEQ ID NO: 38:

```
;  SEQUENCE CHARACTERISTICS:
;    LENGTH: 14 amino acids
;    TYPE: amino acid
;    STRANDEDNESS: single
;    TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-721-458B-38:

Query Match           100.0%; Score 18; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 99;
Matches    4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy    1 KAAK 4
        ||||
Db    6 KAAK 9

RESULT 14
US-08-311-611A-24
; Sequence 24, Application US/08311611A
; Patent No. 5523288
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Jonathan
;  APPLICANT: Kung, Ada H.C.
;  APPLICANT: Lambert, Jr., Lewis H.
;  TITLE OF INVENTION: Method for Treating Gram-Negative Bacterial
;  TITLE OF INVENTION: Infection by Administration of
;  TITLE OF INVENTION: Bactericidal/Permeability-increasing
;  TITLE OF INVENTION:
;  NUMBER OF SEQUENCES: 227
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower, 233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: USA
;   ZIP: 60606-6402
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/311,611A
;   FILING DATE:
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/273,401
;   FILING DATE: 11-JUL-1994
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/125,651
;   FILING DATE: 22-SEP-1993
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Sharp, Jeffrey S.
;   REGISTRATION NUMBER: 31,879
;   REFERENCE/DOCKET NUMBER: 32251
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312/474-6300
;   TELEFAX: 312/474-0448
;   TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 24:
;  SEQUENCE CHARACTERISTICS:
;    LENGTH: 15 amino acids
;    TYPE: amino acid
;    TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   FEATURE:
;    NAME/KEY: misc_feature
;    OTHER INFORMATION: "BPI.23"
US-08-311-611A-24

Query Match           100.0%; Score 18; DB 1; Length 15;
```

```
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches    4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy    1 KAAK 4
        ||||
Db    8 KAAK 11

RESULT 15
US-08-311-611A-115
; Sequence 115, Application US/08311611A
; Patent No. 5523288
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Jonathan
;  APPLICANT: Kung, Ada H.C.
;  APPLICANT: Lambert, Jr., Lewis H.
;  TITLE OF INVENTION: Method for Treating Gram-Negative Bacterial
;  TITLE OF INVENTION: Infection by Administration of
;  TITLE OF INVENTION: Bactericidal/Permeability-increasing
;  TITLE OF INVENTION:
;  NUMBER OF SEQUENCES: 227
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower, 233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: USA
;   ZIP: 60606-6402
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/311,611A
;   FILING DATE:
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/273,401
;   FILING DATE: 11-JUL-1994
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/125,651
;   FILING DATE: 22-SEP-1993
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Sharp, Jeffrey S.
;   REGISTRATION NUMBER: 31,879
;   REFERENCE/DOCKET NUMBER: 32251
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312/474-6300
;   TELEFAX: 312/474-0448
;   TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 115:
;  SEQUENCE CHARACTERISTICS:
;    LENGTH: 15 amino acids
;    TYPE: amino acid
;    TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   FEATURE:
;    NAME/KEY: misc_feature
;    OTHER INFORMATION: "BPI.119"
;   FEATURE:
;    NAME/KEY: Modified-site
;    LOCATION: 7
;    OTHER INFORMATION: /label= Substituted-Ala
;    OTHER INFORMATION: /note= "The alanine at position 7 is beta-1-
;    OTHER INFORMATION: naphthyl-substituted."
;   FEATURE:
;    NAME/KEY: Modified-site
;    LOCATION: 10
;    OTHER INFORMATION: /label= Substituted-Ala
;    OTHER INFORMATION: /note= "The alanine at position 10 is beta-1-
;    OTHER INFORMATION: naphthyl-substituted."
US-08-311-611A-115
```

Query Match   100.0%; Score 18; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
        ||||
Db   8 KAAK 11

Search completed: April 24, 2001, 16:36:23
Job time: 303 sec

OM protein - protein search, using sw model

Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5

Run on: April 24, 2001. 16:41:54 ; Search time 74.56 Seconds
(without alignments)
3.687 Million cell updates/sec

Title: US-09-340-736-3
Perfect score: 18
Sequence: 1 KAAK 4

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 58722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 1000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	DB ID	Description
1	18	100.0	22	2	D47256	kinetoplast DNA-as
2	18	100.0	26	2	A28108	atrial gland pepti
3	18	100.0	27	2	S05454	histone H2A - brin
4	18	100.0	33	2	S61846	hrpY protein - Pse
5	18	100.0	35	2	B45316	sperm-specific pro
6	18	100.0	36	2	F69800	hypothetical prote
7	18	100.0	38	2	PH1920	annexin-like 40K p
8	18	100.0	39	2	A45479	GTP-binding regula
9	18	100.0	45	2	S04941	protamine phi-3.1
10	18	100.0	45	2	S10544	protamine phi-3.2
11	18	100.0	45	2	S10545	protamine phi-3.3
12	18	100.0	45	2	A34639	pectin lyase (EC 4
13	18	100.0	45	2	B43859	ORF overlapping ca
14	18	100.0	54	2	S78292	ribosomal protein
15	18	100.0	54	2	T44359	hypothetical prote
16	18	100.0	58	2	S78240	ribosomal protein
17	18	100.0	58	2	F72128	ribosomal protein
18	18	100.0	61	2	H75184	protein translocas
19	18	100.0	64	2	D69350	hypothetical prote
20	18	100.0	64	2	T22415	hypothetical prote
21	18	100.0	67	2	C75164	archaeal histone a
22	18	100.0	69	2	S66650	proteinase inhibit
23	18	100.0	70	2	H71177	probable archaeal
24	18	100.0	72	2	G42512	G-ORF-3 protein -
25	18	100.0	72	2	A81223	hypothetical prote
26	18	100.0	73	2	F64031	hypothetical prote
27	18	100.0	73	2	F22571	hypothetical prote
28	18	100.0	78	2	S06918	DNA-binding protei
29	18	100.0	78	2	C19434	hypothetical prote
30	18	100.0	79	2	D81188	hypothetical prote
31	18	100.0	84	2	I58177	adrenergic recepto
32	18	100.0	85	2	T08484	regulatory protein
33	18	100.0	86	2	T10047	ribosomal protein
34	18	100.0	88	1	A36614	actobindin - Acant
35	18	100.0	88	2	A69508	LSU ribosomal prot
36	18	100.0	90	2	A60526	complement C3 - ax
37	18	100.0	92	2	A75410	ribosomal protein
38	18	100.0	92	2	E64782	hypothetical prote
39	18	100.0	92	2	S34115	sperm-specific pro
40	18	100.0	94	2	T24826	hypothetical prote
41	18	100.0	95	2	H71711	ribosomal protein
42	18	100.0	96	2	B71528	hypothetical prote
43	18	100.0	96	2	B81682	conserved hypothet
44	18	100.0	96	2	S60889	hypothetical prote
45	18	100.0	97	2	B71719	ribosomal protein

ALIGNMENTS

RESULT 1
D47256
kinetoplast DNA-associated protein p15 - Crithidia fasciculata (fragment)
C;Species: Crithidia fasciculata
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
R;Xu, C.; Ray, D.S.
Proc. Natl. Acad. Sci. U.S.A. 90, 1786-1789, 1993
A;Title: Isolation of proteins associated with kinetoplast DNA networks in vivo.
A;Reference number: A47256; MUID:93189582
A;Accession: D47256
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-22 <XDI>
A;Note: sequence extracted from NCBI backbone (NCBIP:126909)

Query Match 100.0%; Score 18; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAK 4
       ||||
Db 13 KAAK 16

RESULT 2
A28108
atrial gland peptide D1 - California sea hare (fragment)
C;Species: Aplysia californica (California sea hare)
C;Date: 28-Aug-1989 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: A28108
R;Nagle, G.T.; Painter, S.D.; Blankenship, J.E.; Kurosky, A.
J. Biol. Chem. 263, 9223-9237, 1988
A;Title: Proteolytic processing of egg-laying hormone-related precursors in Aplysia.
A;Reference number: A92727; MUID:88243802
A;Accession: A28108
A;Molecule type: protein
A;Residues: 1-26 <NAG>

Query Match 100.0%; Score 18; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAK 4
       ||||
Db 17 KAAK 20

RESULT 3
S05454

histone H2A - brine shrimp (fragment)
C;Species: Artemia sp. (brine shrimp)
C;Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 20-Jun-2000
R;Cruces, J.; Diaz-Guerra, M.; Gil, I.; Renart, J.
Nucleic Acids Res. 17, 6283-6297, 1989
A;Title: The 5S rRNA-histone repeat in the crustacean Artemia: structure, polymorphism a
A;Reference number: S05454; MUID:89366657
A;Accession: S05454
A;Status: translation not shown
A;Molecule type: DNA
A;Residues:1-27 <CRU>
A;Cross-references: EMBL:X14816; NID:g5695; PIDN:CAA32921.1; PID:g1335605
C;Superfamily: histone H2A

Query Match   100.0%; Score 18; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 KAAK 4
        ||||
Db   23 KAAK 26

RESULT 4
S61846
hrpY protein - Pseudomonas solanacearum
C;Species: Pseudomonas solanacearum
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C;Accession: S61846
R;van Gijsegem, F.; Gough, C.; Zischek, C.; Niqueux, E.; Arlat, M.; Genin, S.; Barberis,
Mol. Microbiol. 15, 1095-1114, 1995
A;Title: The hrp gene locus of Pseudomonas solanacearum, which controls the production o
is complex.
A;Reference number: S61846; MUID:95349395
A;Accession: S61846
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues:1-33 <VAN>
A;Cross-references: EMBL:Z14056; NID:g550397; PID:g550398
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1994
C;Genetics:
A;Gene: hrpY

Query Match   100.0%; Score 18; DB 2; Length 33;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 KAAK 4
        ||||
Db   26 KAAK 29

RESULT 5
B45316
sperm-specific protamine-like protein - California mussel (fragment)
N;Alternate names: phil; PL-III
C;Species: Mytilus californianus (California mussel)
C;Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 16-Feb-1997
C;Accession: B45316
R;Carlos, S.; Jutglar, L.; Borrell, I.; Hunt, D.F.; Ausio, J.
J. Biol. Chem. 268, 185-194, 1993
A;Title: Sequence and characterization of a sperm-specific histone H1-like protein of My
A;Reference number: A45316; MUID:93106999
A;Accession: B45316
A;Status: preliminary
A;Molecule type: protein
A;Residues:1-35 <CAR>
A;Note: sequence extracted from NCBI backbone (NCBIP:121400)
C;Keywords: DNA binding

Query Match   100.0%; Score 18; DB 2; Length 35;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 KAAK 4
        ||||
Db   24 KAAK 27

RESULT 6
F69800
hypothetical protein yfhE - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: F69800
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
A.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fujita, M.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A65580; MUID:98044033
A;Accession: F69800
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues:1-36 <KUN>
A;Cross-references: GB:Z99108; GB:AL009125; NID:g2633055; PIDN:CAB12679.1; PID:e11828
A;Experimental source: strain 168
C;Genetics:
A;Gene: yfhE

Query Match   100.0%; Score 18; DB 2; Length 36;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 KAAK 4
        ||||
Db   31 KAAK 34

RESULT 7
PH1920
annexin-like 40K protein - sooty sea hare (fragment)
N;Alternate names: lipocortin
C;Species: Aplysia brasiliana (sooty sea hare)
C;Date: 10-Mar-1994 #sequence_revision 28-Oct-1994 #text_change 17-Mar-1999
C;Accession: PH1920
R;Raju, U.; Nunez-Regueiro, M.; Cook, R.; Kaetzel, M.A.; Yeung, S.C.J.; Eskin, A.
J. Neurochem. 61, 1236-1245, 1993
A;Title: Identification of an annexin-like protein and its possible role in the Aplys
A;Reference number: PH1920; MUID:93389472
A;Accession: PH1920
A;Molecule type: protein
A;Residues:1-38 <RAJ>
A;Experimental source: eye
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: calcium binding

Query Match   100.0%; Score 18; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 KAAK  4
          ||||
Db     13 KAAK 16

RESULT  8
A45479
GTP-binding regulatory protein Gi alpha-3 chain - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
R;Holtzman, E.J.; Kinane, T.B.; West, K.; Soper, B.W.; Ausiello, D.A.; Ercola
J.;Biol. Chem. 268, 3964-3975, 1993
A;Title: Transcriptional regulation of G-protein alpha i subunit genes in LLC-PKI renal
A;Reference number: A45479; MUID:93179393
A;Accession: A45479
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-39 <HOL>
A;Cross-references: GB:CJ0731.7; NID:g164483; PIDN:AAA31049.1; PID:g552365
A;Experimental source: LLC-PKI renal epithelial cells
A;Note: sequence extracted from NCBI backbone (NCBIN:125772, NCBIP:125774)
C;Superfamily: GTP-binding regulatory protein Gs alpha chain
C;Keywords: blocked amino end; lipoprotein; myristylation; thiolester bond
F;2/Modified site: myristylated amino end (GLY) (in mature form) #status predicted
F;3/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match      100.0%;  Score 18;  DB 2;  Length 39;
Best Local Similarity  100.0%;  Pred. No. 2.4e+02;
Matches  4;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 KAAK  4
          ||||
Db     29 KAAK 32

RESULT  9
S04941
protamine phi-3.1 - California mussel
N;Alternate names: protamine PL-IV.1
C;Species: Mytilus californianus (California mussel)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 11-Jan-2000
C;Accession: S04941
R;Ausio, J.; McParland, R.
J.;Biochem. 182, 569-576, 1989
A;Title: Sequence and characterization of the sperm-specific protein phi3 from Mytilus c
A;Reference number: S04941; MUID:89325302
A;Accession: S04941
A;Molecule type: protein
A;Residues: 1-45 <AUS>
A;Note: 6-Thr, 17-Thr, and 18-Thr was also found
C;Superfamily: histone H1
C;Keywords: DNA binding; nucleus

Query Match      100.0%;  Score 18;  DB 2;  Length 45;
Best Local Similarity  100.0%;  Pred. No. 2.7e+02;
Matches  4;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 KAAK  4
          ||||
Db     32 KAAK 35

RESULT  10
S10544
protamine phi-3.2 - California mussel
N;Alternate names: protamine PL-IV.2
C;Species: Mytilus californianus (California mussel)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 11-Jan-2000
C;Accession: S10544
```

```
R;Ausio, J.; McParland, R.
Eur.J. Biochem. 182, 569-576, 1989
A;Title: Sequence and characterization of the sperm-specific protein phi3 from Mytilu
A;Reference number: S04941; MUID:89325302
A;Accession: S10544
A;Molecule type: protein
A;Residues: 1-45 <AUS>
C;Superfamily: histone H1
C;Keywords: DNA binding; nucleus

Query Match      100.0%;  Score 18;  DB 2;  Length 45;
Best Local Similarity  100.0%;  Pred. No. 2.7e+02;
Matches  4;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 KAAK  4
          ||||
Db     32 KAAK 35

RESULT  11
S10545
protamine phi-3.3 - California mussel
N;Alternate names: protamine PL-IV.3
C;Species: Mytilus californianus (California mussel)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 11-Jan-2000
C;Accession: S10545
R;Ausio, J.; McParland, R.
Eur. J. Biochem. 182, 569-576, 1989
A;Title: Sequence and characterization of the sperm-specific protein phi3 from Mytilu
A;Reference number: S04941; MUID:89325302
A;Accession: S10545
A;Molecule type: protein
A;Residues: 1-45 <AUS>
C;Superfamily: histone H1
C;Keywords: DNA binding; nucleus

Query Match      100.0%;  Score 18;  DB 2;  Length 45;
Best Local Similarity  100.0%;  Pred. No. 2.7e+02;
Matches  4;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 KAAK  4
          ||||
Db     32 KAAK 35

RESULT  12
A34639
pectin lyase (EC 4.2.2.10) - Erwinia carotovora (fragment)
C;Species: Erwinia carotovora
C;Date: 13-Jul-1990 #sequence_revision 13-Jul-1990 #text_change 08-Oct-1999
C;Accession: A34639
R;Nishida, T.; Suzuki, T.; Ito, K.; Kanio, Y.; Izaki, K.
Biochem. Biophys. Res. Commun. 168, 801-808, 1990
A;Title: Cloning and expression of pectin lyase gene from Erwinia carotovora in Esche
A;Reference number: A34639; MUID:90241256
A;Accession: A34639
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-45 <NIS>
A;Cross-references: GB:M35271; NID:g148473; PIDN:AAA24855.1; PID:g551856
C;Keywords: carbon-oxygen lyase

Query Match      100.0%;  Score 18;  DB 2;  Length 45;
Best Local Similarity  100.0%;  Pred. No. 2.7e+02;
Matches  4;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 KAAK  4
          ||||
Db     16 KAAK 19
```

RESULT 13
B43859
ORF overlapping carboxyl terminus of IpaA - Shigella dysenteriae (fragment)
C;Species: Shigella dysenteriae
C;Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
C;Accession: B43859
R;Yao. R.; Palchaudhuri, S.
A;Title: Nucleotide sequence and transcriptional regulation of a positive regulatory gen
Infect. Immun. 60, 1163-1169, 1992
A;Reference number: A43859; MUID:92175961
A;Contents: CG097, large invasion plasmid
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-45 <IAO>
A;Note: sequence extracted from NCBI backbone (NCBIN:85367, NCBIP:85368)

Query Match   100.0%; Score 18; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
            ||||
Db   26 KAAK 29

RESULT 14
S78292
ribosomal protein L32', chloroplast - Odontella sinensis chloroplast
C;Species: chloroplast Odontella sinensis
C;Date: 17-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C;Accession: S78292
R;Kowallik, K.V.; Stoebe, B.; Schaffran, I.; Kroth-Pancic, P.; Freier, U.
Plant Mol. Biol. Rep. 13, 336-342, 1995
A;Title: The Chloroplast Genome of a chlorophyll a+c- containing Alga, Odontella sinensi
A;Reference number: S78238
A;Accession: S78292
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-54 <KOW>
A;Cross-references: EMBL:Z67753; NID:g1185127; PIDN:CAA91665.1; PID:g1185182
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995
C;Genetics:
A;Gene: rpl32'
A;Genome: chloroplast
C;Superfamily: rice chloroplast ribosomal protein L32
C;Keywords: chloroplast; protein biosynthesis; ribosome

Query Match   100.0%; Score 18; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
            ||||
Db   25 KAAK 29

RESULT 15
T44359
hypothetical protein ruvB [imported] - Clostridium histolyticum (fragment)
C;Species: Clostridium histolyticum
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Sep-2000
C;Accession: T44359
R;Matsushita, O.; Jung, C.M.; Katayama, S.; Minami, J.; Takahashi, Y.; Okabe, A.
J. Bacteriol. 181, 923-933, 1999
A;Title: Gene duplication and multiplicity of collagenases in Clostridium histolyticum.
A;Reference number: Z22752; MUID:99121032
A;Accession: T44359
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA

A;Residues: 1-54 <MAT>
A;Cross-references: EMBL:AB014075; NID:g3868863; PIDN:BAA34546.1; PID:g3892650
A;Experimental source: strain JCM 1403
C;Genetics:
A;Note: ruvB
C;Superfamily: ruvB protein

Query Match   100.0%; Score 18; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
            ||||
Db   44 KAAK 47

Search completed: April 24, 2001, 16:41:56
Job time: 461 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:42:45 ; Search time 44.88 Seconds
                                         (without alignments)
                                         3.053 Million cell updates/sec

Title: US-09-340-736-3
Perfect score: 18
Sequence: 1 KAAK 4

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

?ched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	18	100.0	35	1	PHI1_MYTCA	P35422 mytilus cal
2	18	100.0	45	1	PHI3_MYTCA	P11860 mytilus cal
3	18	100.0	57	1	RK32_ODOSI	P49564 odontella s
4	18	100.0	58	1	RS2l_CHLPN	Q9z9f0 chlamydia p
5	18	100.0	67	1	HPHB_PYRHO	O74092 pyrococcus
6	18	100.0	69	1	Y4LI_RHISN	P55549 rhizobium s
7	18	100.0	72	1	YVGB_VACCC	P20566 vaccinia vi
8	18	100.0	73	1	YE85_HAEIN	P44212 haemophilus
9	18	100.0	78	1	PHIO_HOLTU	P14309 holothuria
10	18	100.0	82	1	DNAJ_STRAL	P96457 streptomyce
11	18	100.0	84	1	A1AA_MOUSE	P97718 mus musculu
12	18	100.0	85	1	KOCl_ECOLI	Q52277 escherichia
13	18	100.0	86	1	RS20_MYCLE	O33132 mycobacteri
14	18	100.0	88	1	ACTO_ACACA	P18281 acanthamoeb
15	18	100.0	88	1	RL3l_ARCFU	O28213 archaeoglob
16	18	100.0	91	1	PHI1_MYTED	O04621 mytilus edu
17	18	100.0	92	1	YBBV_ECOLI	P75711 escherichia
18	18	100.0	93	1	RS19_LEPIN	O9xd32 leptospira
19	18	100.0	95	1	RS18_RICPR	O9zea5 rickettsia
20	18	100.0	96	1	Y335_CHIFR	O84339 chlamydia t
21	18	100.0	96	1	RL28_RICPR	O9p55B rickettsia
22	18	100.0	97	1	YORE_PYRHO	O9ze48 pyrococcus
23	18	100.0	99	1	ELS_SHEEP	P11537 ovis aries
24	18	100.0	100	1	GlUC_RANCA	P15438 rana catesb
25	18	100.0	103	1	Hl1_BOVIN	P02253 bos taurus
26	18	100.0	104	1	RS25_YEAST	P07282 saccharomyc
27	18	100.0	108	1	SSI_STRON	P80597 streptovert
28	18	100.0	109	1	GLR2_SCHPO	O9ut12 schizosacch
29	18	100.0	110	1	RLAl_ALTAL	P49148 alternaria
30	18	100.0	110	1	VFUS_VACCV	Pl1258 vaccinia vi
31	18	100.0	110	1	VFUS_VARV	P33816 variola vir
32	18	100.0	112	1	HMGD_DROME	Q05783 drosophila
34	18	100.0	112	1	POR3_PIG	Q29380 sus scrofa
35	18	100.0	113	1	CYC7_YEAST	P00045 saccharomyc
36	18	100.0	114	1	HMGA_CHITE	P40622 chironomus
37	18	100.0	114	1	KCGL_BACSU	P19638 bacillus su
38	18	100.0	116	1	RL19_STRCO	O69883 streptomyco
39	18	100.0	116	1	RL20_HAEIN	P44358 haemophilus
40	18	100.0	117	1	Y743_PYRAB	Q9uzm5 pyrococcus
41	18	100.0	118	1	RL20_THEMA	Q9xls8 thermotoga
42	18	100.0	118	1	H1B_PLADU	P06895 platynereis
43	18	100.0	119	1	RL19_MYCPN	P75133 mycoplasma
44	18	100.0	120	1	YG9_YEAST	P53188 saccharomyc
45	18	100.0	121	1	H1A_PLADU	P06894 platynereis

ALIGNMENTS

RESULT   1
PHI1_MYTCA
ID   PHI1_MYTCA   STANDARD;   PRT;   35 AA.
AC   P35422;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   SPERM-SPECIFIC PROTEIN PHI-1 (PL-III) (SPERM-SPECIFIC PROTAMINE-LIKE
DE   PROTEIN) (FRAGMENT).
OS   Mytilus californianus (California mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
OX   NCBI_TaxID=6549;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Sperm;
RX   MEDLINE=93106999; PubMed=7677995;
RA   Carlos S., Jutglar L., Borrell I., Hunt D.F., Ausio J.;
RT   "Sequence and characterization of a sperm-specific histone H1-like
RT   protein of Mytilus californianus.";
RL   J. Biol. Chem. 268:185-194(1993).
CC   -!- FUNCTION: INVOLVED IN NUCLEAR BASIC PROTEIN TRANSITION: HISTONES
CC   ARE REPLACED BY SPERMATID SPECIFIC PROTEINS WHICH ARE THEMSELVES
CC   REPLACED BY PROTAMINES IN LATE SPERMATIDS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: SPERM.
DR   PIR; B45316; B45316.
KW   Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW   Nuclear protein.
FT   NON_TER   35   35
SQ   SEQUENCE   35 AA; 3759 MW; 97459D27A76F4BD1 CRC64;

     Query Match   100.0%; Score 18; DB 1; Length 35;
     Best Local Similarity   100.0%; Fred. No. 81;
     Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
            ||||
Db   24 KAAK 27

RESULT   2
PHI3_MYTCA
ID   PHI3_MYTCA   STANDARD;   PRT;   45 AA.
AC   P11860;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-AUG-1990 (Rel. 15, Last annotation update)
DE   SPERM-SPECIFIC PROTEIN PHI-3 (PL-IV).
OS   Mytilus californianus (California mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
OX   NCBI_TaxID=6549;
RN   [1]
RP   SEQUENCE.

RX   MEDLINE=89325302; PubMed=2666130;
RA   Auslo J., McParland R.;
RT   "Sequence and characterization of the sperm-specific protein phi 3
RT   from Mytilus californianus.";
RL   Eur. J. Biochem. 182:569-576(1989).
CC   -!- FUNCTION: INVOLVED IN NUCLEAR BASIC PROTEIN TRANSITION: HISTONES
CC   ARE REPLACED BY SPERMATID SPECIFIC PROTEINS WHICH ARE THEMSELVES
CC   REPLACED BY PROTAMINES IN LATE SPERMATIDS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- MISCELLANEOUS: THE SEQUENCE OF COMPONENT I IS SHOWN.
DR   PIR; S04941; S04941.
DR   PIR; S10544; S10544.
DR   PIR; S10545; S10545.
KW   Chromosomal protein; Nucleosome core: Spermatogenesis; DNA-binding;
KW   Nuclear protein.
FT   VARIANT   12   12   A -> T (IN COMPONENT II).
FT   VARIANT   39   39   K -> R (IN COMPONENT III).
FT   VARIANT   6   6   S -> T (POSSIBLE VARIANT OF COMPONENT I).
FT   VARIANT   17   17   S -> T (POSSIBLE VARIANT OF COMPONENT I).
FT   VARIANT   18   18   S -> T (POSSIBLE VARIANT OF COMPONENT I).
SQ   SEQUENCE   45 AA; 4912 MW; 78A68DF2CH77C4E1 CRC64;

Query Match   100.0%; Score 18; DB 1; Length 45;
Best Local Similarity   100.0%; Pred. No. 1e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
            |||
Db   32 KAAK 35

RESULT 3
RK32_ODOSI   STANDARD;   PRT;   57 AA.
AC   P49564;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   CHLOROPLAST 50S RIBOSOMAL PROTEIN L32.
GN   RPL32 AND RPL32'.
OS   Odontella sinensis.
OG   Chloroplast.
OC   Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC   Biddulphiophycidae; Eupodiscales; Eupodiscaceae; Odontella.
OX   NCBI_TaxID=2839;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kowallik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
RT   "The chloroplast genome of a chlorophyll a+c-containing alga,
RT   Odontella sinensis.";
RL   Plant Mol. Biol. Rep. 13:336-342(1995).
-!- SIMILARITY: BELONGS TO THE L32P FAMILY OF RIBOSOMAL PROTEINS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Z67753; CAA91613.1; -.
DR   EMBL; Z67753; CAA91665.1; -.
DR   Mendel; 4723; ODOSi;rpl32l.
DR   InterPro; IPR002677; -.
DR   Pfam; PF01783; Ribosomal_L32p; 1.
KW   Ribosomal protein; Chloroplast.
FT   INIT_MET   0   0   BY SIMILARITY.
SQ   SEQUENCE   57 AA; 6374 MW; 5253B73505AFA3C0 CRC64;

Query Match   100.0%; Score 18; DB 1; Length 57;

Best Local Similarity   100.0%; Pred. No. 1.2e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
            |||
Db   25 KAAK 28

RESULT 4
RS21_CHLPN
ID   RS21_CHLPN   STANDARD;   PRT;   58 AA.
AC   Q9Z9F0; Q8J0X5;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   30S RIBOSOMAL PROTEIN S21.
GN   RPSU OR RS21 OR CPN0031 OR CP0745.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
CC   -!- SIMILARITY: BELONGS TO THE S21P FAMILY OF RIBOSOMAL PROTEINS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AE001588; AAD18184.1; -.
DR   EMBL; AE002233; AAF38550.1; -.
DR   EMBL; AP002545; BAA98243.1; -.
DR   TIGR; CP0745; -.
DR   InterPro; IPR001911; -.
DR   Pfam; PF01165; Ribosoma_S21; 1.
DR   PRINTS; PR00976; RIBOSOMALS21.
DR   PROSITE; PS01181; RIBOSOMAL_S21; FALSE_NEG.
KW   Ribosomal protein.
SQ   SEQUENCE   58 AA; 6635 MW; 74118AE25600177 CRC64;

Query Match   100.0%; Score 18; DB 1; Length 58;
Best Local Similarity   100.0%; Pred. No. 1.3e+02;

```
Matches    4; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

Qy     1 KAAK 4
         ||||
Db    30 KAAK 33

RESULT 5
HPHB_PYRHO
ID   HPHB_PYRHO          STANDARD;          PRT;         67 AA.
AC   O74092;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PROBABLE ARCHAEAL HISTONE B.
GN   PHS046.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.;
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- FUNCTION: BINDS AND COMPACT DNA (95 TO 150 BASE PAIRS) TO FORM
CC       NUCLEOSOME-LIKE STRUCTURES THAT CONTAIN POSITIVE DNA SUPERCOILS
CC       INCREASES THE RESISTANCE OF DNA TO THERMAL DENATURATION (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HMF FAMILY OF ARCHAEAL HISTONES.
CC   ------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------
DR   EMBL; AP000007; BAA30815.1; ALT_INIT.
DR   HSSP; P48781; 1HTA.
DR   InterPro; IPR000947; -.
DR   Pfam; PF00808; CBFD_NFYB_HMF; 1.
KW   DNA-binding; Multigene family.
SQ   SEQUENCE  67 AA;  7366 MW;  00A04ACE5FAF54A CRC64;

Query Match          100.0%;  Score 18;  DB 1;  Length 67;
Best Local Similarity 100.0%;  Pred. No. 1.4e+02;
Matches    4; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

Qy     1 KAAK 4
         ||||
Db    24 KAAK 27

RESULT 6
Y4LI_RHISN
ID   Y4LI_RHISN          STANDARD;          PRT;         69 AA.
AC   P55549;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPOTHETICAL 7.3 KDA PROTEIN Y4LI.
GN   Y4LI
OS   Rhizobium sp. (strain NGR234).
```

```
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=394;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97305956; PubMed=9163424;
RA   Freiberg C.A., Feilay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA   Perret X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
RL   Nature 387:394-401(1997).
CC   -!- SIMILARITY: NONE OBVIOUS.
CC   ------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------
DR   EMBL; AE000083; AAB91761.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE  69 AA;  7275 MW;  2349316222798BDF CRC64;

Query Match          100.0%;  Score 18;  DB 1;  Length 69;
Best Local Similarity 100.0%;  Pred. No. 1.5e+02;
Matches    4; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

Qy     1 KAAK 4
         ||||
Db    25 KAAK 28

RESULT 7
YVGB_VACCC
ID   YVGB_VACCC          STANDARD;          PRT;         72 AA.
AC   P20566;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   HYPOTHETICAL 8.4 KDA PROTEIN.
GN   ORF B.
OS   Vaccinia virus (strain Copenhagen).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   orthopoxvirus.
OX   NCBI_TaxID=10249;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91021027; PubMed=2219722;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [2]
RP   COMPLETE GENOME.
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RL   Virology 179:517-563(1990).
CC   ------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------
DR   EMBL; M35027; AAA48073.1; -.
DR   PIR; G42512; G42512
KW   Hypothetical protein.
SQ   SEQUENCE  72 AA;  8369 MW;  2E8597B39302728B CRC64;
```

Query Match   100.0%; Score 18; DB 1; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;
Matches   4; Conservative   0; Mismatches   0;

```
Qy      1 KAAK 4
          ||||
Db     10 KAAK 13
```

RESULT 8
ID   YE85_HAEIN   STANDARD;   PRT;   73 AA.
AC   P44212;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   VERY HYPOTHETICAL PROTEIN HI1485 IN MU-LIKE PROPHAGE FLUMU REGION.
GN   HI1485.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U32826; AAC23140.1; -.
     TIGR; HI1485; -.
DE   Hypothetical protein.
SQ   SEQUENCE   73 AA; 8834 MW; AC20D81B670923FD CRC64;

Query Match   100.0%; Score 18; DB 1; Length 73;
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;
Matches   4; Conservative   0; Mismatches   0;

```
Qy      1 KAAK 4
          ||||
Db     23 KAAK 26
```

RESULT 9
ID   PHIO_HOLTU   STANDARD;   PRT;   78 AA.
AC   P14309;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-APR-1990 (Rel. 14, Last annotation update)
DE   SPERM-SPECIFIC PROTEIN PHI-0.
OS   Holothuria tubulosa (Sea cucumber).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Holothuroidea; Aspidochirotacea; Aspidochirotida; Holothuriidae;
OC   Holothuria.
OX   NCBI_TaxID=7685;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Sperm;
RX   MEDLINE=90098774; PubMed=2602115;
RA   Prats E., Cornudella L., Ruiz-Carrillo A.;
RT   "Nucleotide sequence of a cDNA for phi, a histone to protamine
RT   transition protein from sea cucumber spermatozoa.";
RL   Nucleic Acids Res. 17:10097-10097(1989).
CC   -!- FUNCTION: INVOLVED IN NUCLEAR BASIC PROTEIN TRANSITION: HISTONES
CC   ARE REPLACED BY SPERMATID SPECIFIC PROTEINS WHICH ARE THEMSELVES
CC   REPLACED BY PROTAMINES IN LATE SPERMATIDS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M27408; AAA29204.1; -.
DR   EMBL; X16364; CAA34410.1; -.
DR   PIR; S06918; S06918.
KW   Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW   Nuclear protein.
SQ   SEQUENCE   78 AA; 8680 MW; ABFA27075961 4F2C CRC64;

Query Match   100.0%; Score 18; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;
Matches   4; Conservative   0; Mismatches   0;

```
Qy      1 KAAK 4
          ||||
Db     56 KAAK 59
```

RESULT 10
ID   DNAJ_STRAL   STANDARD;   PRT;   82 AA.
AC   P96457;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   CHAPERONE PROTEIN DNAJ (FRAGMENT).
GN   DNAJ.
OS   Streptomyces albus G.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1962;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J1074;
RX   MEDLINE=97157943; PubMed=9004222;
RA   Grandvalet C., Servant P., Mazodier P.;
RT   "Disruption of hspR, the repressor gene of the dnaK operon in
RT   Streptomyces albus G.";
RL   Mol. Microbiol. 23:77-84(1997).
CC   -!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,
CC   THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).
CC   -!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way

CC　modified and this statement is not removed. Usage by and for commercial
CC　entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC　or send an email to license@isb-sib.ch).
CC　-!-
DR　EMBL; U43299; AAB48077.1; -.
DR　InterPro; IPR001623;
DR　PROSITE; PS00636; DNAJ_1; PARTIAL.
DR　PROSITE; PS50076; DNAJ_2; PARTIAL.
DR　PROSITE; PS00637; DNAJ_CXXCXGXG; PARTIAL.
KW　Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.
FT　NON_TER　　1　　1
SQ　SEQUENCE　　82 AA;　8663 MW;　BFE22CCF2845F5349 CRC64;

Query Match　　　　100.0%;　Score 18;　DB 1;　Length 82;
Best Local Similarity　100.0%;　Pred. No. 1.7e+02;
Matches　　4;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 KAAK 4
　　　　|||
Db　77 KAAK 80

RESULT 11
A1AA_MOUSE
ID　A1AA_MOUSE　　STANDARD;　　PRT;　　　84 AA.
AC　P97718;
DT　15-JUL-1998 (Rel. 36, Created)
DT　15-JUL-1998 (Rel. 36, Last sequence update)
DT　15-JUL-1999 (Rel. 38, Last annotation update)
DE　ALPHA-1A ADRENERGIC RECEPTOR (ALPHA 1A-ADRENOCEPTOR) (ALPHA-1C
DE　ADRENERGIC RECEPTOR) (FRAGMENT).
DE　ADRA1A OR ADRA1C.
GN　Mus musculus (Mouse).
OS　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC　Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX　NCBI_TaxID=10090;
RN　[1]
RP　SEQUENCE FROM N.A.
RC　TISSUE=Brain;
RX　MEDLINE=96064818; PubMed=7595553;
RA　Alonso-Llamazares A., Zamanillo D., Casanova E., Ovalle S., Calvo P.,
RA　Chinchetru M.A.;
RT　"Molecular cloning of alpha 1d-adrenergic receptor and tissue
RT　distribution of three alpha 1-adrenergic receptor subtypes in
RT　mouse.";
RL　J. Neurochem. 65:2387-2392(1995).
CC　-!- FUNCTION: THIS ALPHA-ADRENERGIC RECEPTOR MEDIATES ITS ACTION BY
CC　ASSOCIATION WITH G PROTEINS THAT ACTIVATE A PHOSPHATIDYLINOSITOL-
CC　CALCIUM SECOND MESSENGER SYSTEM. ITS EFFECT IS MEDIATED BY G(Q)
CC　AND G(11) PROTEINS.
CC　-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC　-!- PTM: CARBOXYL-TERMINAL SER OR THR RESIDUES MAY BE PHOSPHORYLATED
CC　(BY SIMILARITY).
CC　-!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC　---
CC　This SWISS-PROT entry is copyright. It is produced through a collaboration
CC　between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC　the European Bioinformatics Institute. There are no restrictions on its
CC　use by non-profit institutions as long as its content is in no way
CC　modified and this statement is not removed. Usage by and for commercial
CC　entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC　or send an email to license@isb-sib.ch).
CC　-!-
DR　EMBL; S80220; AAB47044.1; -.
DR　MGD; MGI:104773; Adra1c.
DR　InterPro; IPR000276; -.
DR　Pfam; PF00001; 7tm_1; 1.
KW　G-protein coupled receptor; Transmembrane; Glycoprotein;
KW　Multigene family; Phosphorylation; Lipoprotein; Palmitate.
FT　NON_TER　　<1　　1
FT　TRANSMEM　　1　　9　　5 (POTENTIAL).

FT　DOMAIN　　　10　　77　　CYTOPLASMIC (POTENTIAL).
FT　TRANSMEM　　78　　>84　　6 (POTENTIAL).
FT　MOD_RES　　19　　19　　PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT　NON_TER　　84　　84
SQ　SEQUENCE　　84 AA;　9308 MW;　F2E31D6I9A076425 CRC64;

Query Match　　　　100.0%;　Score 18;　DB 1;　Length 84;
Best Local Similarity　100.0%;　Pred. No. 1.7e+02;
Matches　　4;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 KAAK 4
　　　　|||
Db　73 KAAK 76

RESULT 12
KOC1_ECOLI
ID　KOC1_ECOLI　　STANDARD;　　PRT;　　　85 AA.
AC　Q52277;
DT　01-NOV-1997 (Rel. 35, Created)
DT　01-NOV-1997 (Rel. 35, Last sequence update)
DT　01-OCT-2000 (Rel. 40, Last annotation update)
DE　TRANSCRIPTIONAL REPRESSOR PROTEIN KORC.
GN　KORC.
OS　Escherichia coli.
OG　Plasmid IncP-beta R751.
OC　Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC　Escherichia.
OX　NCBI_TaxID=562;
RN　[1]
RP　SEQUENCE FROM N.A.
RC　STRAIN=K12 / C600;
RX　MEDLINE=95291464; PubMed=7773415;
RA　Thomas C.M., Smith C.A., Ibbotson J.P., Johnston L., Wang N.;
RT　"Evolution of the korA-oriV segment of promiscuous IncP plasmids.";
RL　Microbiology 141:1201-1210(1995).
CC　-!- FUNCTION: ACTS WITH KORA AS COREPRESSOR IN THE CONTROL OF THE KILC
CC　AND KILE OPERONS.
CC　---
CC　This SWISS-PROT entry is copyright. It is produced through a collaboration
CC　between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC　the European Bioinformatics Institute. There are no restrictions on its
CC　use by non-profit institutions as long as its content is in no way
CC　modified and this statement is not removed. Usage by and for commercial
CC　entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC　or send an email to license@isb-sib.ch).
CC　-!-
DR　EMBL; U67194; AAC64428.1; -.
KW　Plasmid; Transcription regulation; Repressor; DNA-binding.
FT　DNA_BIND　　28　　47　　H-T-H MOTIF (POTENTIAL).
SQ　SEQUENCE　　85 AA;　9195 MW;　5DCCF940B3CAC269 CRC64;

Query Match　　　　100.0%;　Score 18;　DB 1;　Length 85;
Best Local Similarity　100.0%;　Pred. No. 1.7e+02;
Matches　　4;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 KAAK 4
　　　　|||
Db　39 KAAK 42

RESULT 13
RS20_MYCLE
ID　RS20_MYCLE　　STANDARD;　　PRT;　　　85 AA.
AC　O33132;
DT　15-JUL-1998 (Rel. 36, Created)
DT　15-JUL-1998 (Rel. 36, Last sequence update)
DT　15-JUL-1998 (Rel. 36, Last annotation update)
DE　30S RIBOSOMAL PROTEIN S20.
GN　RPS? OR MLCL536.06.
OS　Mycobacterium leprae.

```
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seeger K., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE S20P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z99125; CAB16149.1; -.
DR   InterPro; IPR002583; -.
DR   Pfam; PF01649; Ribosomal_S20p; 1.
KW   Ribosomal protein; rRNA-binding.
SQ   SEQUENCE   86 AA;  9609 MW;  13B18F2730582790 CRC64;

Query Match             100.0%;  Score 18;  DB 1;  Length 86;
Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 KAAK 4
         ||||
Db   46 KAAK 49


RESULT 14
ACTO_ACACA
ID   ACTO_ACACA     STANDARD;      PRT;    88 AA.
AC   P18281;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ACTOBINDIN.
OS   Acanthamoeba castellanii (Amoeba).
OC   Eukaryota; Acanthamoebidae; Acanthamoeba.
OX   NCBI_TaxID=5755;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=90330610; PubMed=2376577;
RA   Vandekerckhove J., van Damme J., Vancompernolle K., Bubb M.R.,
RA   Lambooy P.K., Korn E.D.;
RL   "The covalent structure of Acanthamoeba actobindin.";
RL   J. Biol. Chem. 265:12801-12805(1990).
CC   -!- FUNCTION: IS ABLE TO BIND TWO ACTIN MONOMERS AT HIGH
CC       CONCENTRATIONS OF G-ACTIN.
CC   -!- SUBUNIT: MONOMER.
DR   PIR; A36614; A36614.
KW   Actin-binding; Repeat; Methylation; Acetylation.
FT   MOD_RES       1      1       ACETYLATION.
FT   MOD_RES      35     35       METHYLATION (TRI-).
FT   MOD_RES      72     72       METHYLATION (TRI-).
FT   REPEAT       15     47       NEARLY IDENTICAL.
FT   REPEAT       51     84       NEARLY IDENTICAL.
SQ   SEQUENCE   88 AA;  9554 MW;  E81A45C521603DA6 CRC64;

Query Match             100.0%;  Score 18;  DB 1;  Length 88;
Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 KAAK 4
         ||||
Db   84 KAAK 87


RESULT 15
RL31_ARCFU
ID   RL31_ARCFU     STANDARD;      PRT;    88 AA.
AC   O28213;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L31E.
GN   RPL31E OR AF2066.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketcham K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fuji C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   -!- SIMILARITY: BELONGS TO THE L31E FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE000961; AAB89205.1; -.
DR   TIGR; AF2066; -.
DR   InterPro; IPR000054; -.
DR   Pfam; PF01198; Ribosomal_L31e; 1.
DR   PROSITE; PS01144; RIBOSOMAL_L31E; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   88 AA;  10488 MW;  04DB178F487EBF54 CRC64;

Query Match             100.0%;  Score 18;  DB 1;  Length 88;
Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 KAAK 4
         ||||
Db   29 KAAK 32


Search completed: April 24, 2001, 16:42:47
Job time: 453 sec
```

```
                    GenCore version 4.5
              Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:       April 24, 2001, 16:40:27 ; Search time 125.5 seconds
                                          (without alignments)
                                          3.736 Million cell updates/sec

Title:        US-09-340-736-3
Perfect score: 18
Sequence:      1 KAAK 4

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

ched:          374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters:    374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     SPTREMBL_15:*
               1:  sp_archea.*
               2:  sp_bacteria:*
               3:  sp_fungi:*
               4:  sp_human:*
               5:  sp_invertebrate:*
               6:  sp_mammal:*
               7:  sp_mhc:*
               8:  sp_organelle:*
               9:  sp_phage:*
              10:  sp_plant:*
              11:  sp_rodent:*
              12:  sp_unclassified:*
              13:  sp_vertebrate:*
              14:  sp_virus:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                        SUMMARIES

Result          %     Query
No.    Score  Match  Length  DB  ID        Description

 1     18    100.0    11    2   Q9R446    Q9r446 neisseria g
 2     18    100.0    20    6   Q9TRW4    Q9trw4 bos taurus
 3     18    100.0    22    2   Q9ZIF1    Q9zif1 neisseria m
 4     18    100.0    22    5   Q9TWU8    Q9twu8 crithidia f
 5     18    100.0    27    4   Q17096    Q17096 artemia sp.
 6     18    100.0    35   10   Q9SPU2    Q9spu2 arabidopsis
 7     18    100.0    36    2   O31573    O31573 bacillus su
 8     18    100.0    37    2   O85035    O85035 mycoplasma
 9     18    100.0    37    5   O61191    O61191 blepharisma
10     18    100.0    38    5   O61189    O61189 blepharisma
11     18    100.0    38    5   O61190    O61190 blepharisma
12     18    100.0    39    6   Q29047    Q29047 sus scrofa
13     18    100.0    45    2   O47512    O47512 erwinia car
14     18    100.0    54    2   Q9ZNJ5    Q9znj5 clostridium
15     18    100.0    54    5   Q9T314    Q9t314 plasmodium
16     18    100.0    58    4   Q9JGK5    Q9jgk5 chlamydia p
17     18    100.0    58   11   O55067    O55067 meriones un
18     18    100.0    61    1   Q9V2S1    Q9v2s1 pyrococcus
19     18    100.0    64    1   O29454    O29454 archaeoglob
```

```
20  18  100.0   64   5   Q20588   Q20588 caenorhabdi
21  18  100.0   67   1   Q9VIF5   Q9vif5 pyrococcus
22  18  100.0   69   4   Q13626   Q13626 homo sapien
23  18  100.0   69   6   Q28101   Q28101 bos taurus
24  18  100.0   69  10   Q9S9F3   Q9s9f3 fagopyrum e
25  18  100.0   69  10   Q98896   Q98896 fagopyrum e
26  18  100.0   69  10   Q98895   Q98895 fagopyrum e
27  18  100.0   70   2   O53012   O53012 rhizobium e
28  18  100.0   70   2   O14377   O14377 schizosacch
29  18  100.0   72   2   O87036   O87036 vibrio chol
30  18  100.0   72   2   Q9KID6   Q9kid6 neisseria m
31  18  100.0   75   2   O56374   O56374 trichodesmi
32  18  100.0   75   2   O30904   O30904 salmonella
33  18  100.0   75   2   Q9Z8F4   Q9z8f4 salmonella
34  18  100.0   78   2   Q94582   Q94582 holochtria
35  18  100.0   78   5   Q9XG97   Q9xg97 caenorhabdi
36  18  100.0   79   2   Q9KGR6   Q9kgr6 neisseria m
37  18  100.0   79   9   Q9MBT3   Q9mbt3 staphylococ
38  18  100.0   81   5   Q9PIF3   Q9pif3 homo sapien
39  18  100.0   81   5   O18495   O18495 styela clav
40  18  100.0   81   5   O18496   O18496 styela clav
41  18  100.0   81  14   Q92514   Q92514 beet virus
42  18  100.0   82   2   Q52483   Q52483 burkholderi
43  18  100.0   84   6   Q9TTK3   Q9ttk3 canis famil
44  18  100.0   84  11   P70648   P70648 rattus norv
45  18  100.0   86   2   O87326   O87326 helicobacte
```

```
                        ALIGNMENTS

RESULT         1
Q9R445
ID    Q9R446;              PRELIMINARY;      PRT;      11 AA.
AC    Q9R446;
DT    01-MAY-2000 (TrEMBLrel. 13, Created)
DT    01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT    01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE    CARBAMOYL-PHOSPHATE SYNTHASE SUBUNIT A (FRAGMENT).
GN    CARA.
OS    Neisseria gonorrhoeae.
OC    Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX    NCBI_TaxID=485;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=MS11, AND FA1090;
RX    MEDLINE=95291461; PubMed=7773412;
RA    Lawson F.S., Billowes F.M., Dillon J.A.;
RT    "Organization of carbamoyl-phosphate synthase genes in Neisseria
RT    gonorrhoeae includes a large, variable intergenic sequence which is
RT    also present in other Neisseria species.";
RL    Microbiology 141:0-0(0).
RN    [2]
RP    SEQUENCE FROM N.A.
RC    STRAIN=MS11, AND FA1090;
RA    Brinkman F.S.L., Francis F.M., Dillon J.R.;
RT    "Complexity of the variable sequence between the carbamoyl-phosphate
RT    synthase genes of Neisseria species.";
RL    Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AF029363; AAC78453.1; -.
DR    EMBL; AF029362; AAC78452.1; -.
FT    NON_TER        1      1
SQ    SEQUENCE   11 AA;   1178 MW;  0C07A8E3DDD33694 CRC64;

Query Match             100.0%;  Score 18;  DB 2;  Length 11;
Best Local Similarity   100.0%;  Pred. No. 88;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 KAAK 4
           |||
Db     6 KAAK 9
```

RESULT 2
ID   Q9TRW4
AC   Q9TRW4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   25 KDA PROTEIN P25, PEPTIDE F5A.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=91372400; PubMed=1909972;
RA   Takahashi M., Tomizawa K., Ishiguro K., Sato K., Omori A., Sato S.,
RA   Shiratsuchi A., Uchida T., Imahori K.;
RT   "A novel brain-specific 25 kDa protein (p25) is phosphorylated by a
RT   Ser/Thr-Pro Kinase (TPK II) from tau protein kinase fractions.";
RL   FEBS Lett. 289:37-43(1991).
SQ   SEQUENCE 20 AA; 2032 MW; 4AADB849A6416897 CRC64;

PRELIMINARY;   PRT;   20 AA.

Query Match   100.0%; Score 18; DB 5; Length 20;
Best Local Similarity   100.0%; Pred. No. 1.6e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
               ||||
Db   17 KAAK 20

RESULT 3
ID   Q9ZIFI
AC   Q9ZIFI;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CARBAMOYL-PHOSPHATE SYNTHASE SUBUNIT A (FRAGMENT).
GN   CARA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1527;
RX   MEDLINE=95291461; PubMed=7773412;
RA   Lawson F.S., Billowes F.M., Dillon J.A.;
RT   "Organization of carbamoyl-phosphate synthase genes in Neisseria
RT   gonorrhoeae includes a large variable intergenic sequence which is
RT   also present in other Neisseria species.";
RL   Microbiology 141:0-0(0).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1527;
RA   Brinkman F.S.L., Francis F.M., Dillon J.R.;
RT   "Complexity of the variable sequence between the carbamoyl-phosphate
RT   synthase genes of Neisseria species.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF029361; AAC78447.1; -.
FT   NON_TER   1   1
SQ   SEQUENCE 22 AA; 2427 MW; 3445B94405B96AF6 CRC64;

PRELIMINARY;   PRT;   22 AA.

Query Match   100.0%; Score 18; DB 2; Length 22;
Best Local Similarity   100.0%; Pred. No. 1.7e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
               ||||

Db   17 KAAK 20

RESULT 4
ID   Q9PWU8
AC   Q9PWU8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   KINETOPLAST DNA-ASSOCIATED PROTEIN (FRAGMENT).
OS   Crithidia fasciculata.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
OX   NCBI_TaxID=5656;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=93189582; PubMed=8446592;
RA   Xu C., Ray D.S.;
RT   "Isolation of proteins associated with kinetoplast DNA networks in
RT   vivo.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:1786-1789(1993).
SQ   SEQUENCE 22 AA; 1991 MW; 5B1C4941E0E8F9B2 CRC64;

PRELIMINARY;   PRT;   22 AA.

Query Match   100.0%; Score 18; DB 5; Length 22;
Best Local Similarity   100.0%; Pred. No. 1.7e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
               ||||
Db   13 KAAK 16

RESULT 5
ID   Q17096
AC   Q17096;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ARTEMIA 5S RRNA (PARTSH-B), I.
OS   Artemia sp. (Brine shrimp).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda; Anostraca;
OC   Artemiidae; Artemia.
OX   NCBI_TaxID=6662;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PARTHOGENETIC POPULATION FROM EURASIA.;
RX   MEDLINE=89366657; PubMed=2570403;
RA   Cruces J., Diaz-Guerra M., Gil I., Renart J.;
RT   "The 5S rRNA-histone repeat in the crustacean Artemia: structure,
RT   polymorphism and variation of the 5S rRNA segment in different
RT   populations.";
RL   Nucleic Acids Res. 17:6283-6297(1989).
DR   EMBL; X14816; CAA32921.1; -.
KW   DNA-binding.
SQ   SEQUENCE 27 AA; 2759 MW; 0AE2088B586AA63B CRC64;

PRELIMINARY;   PRT;   27 AA.

Query Match   100.0%; Score 18; DB 5; Length 27;
Best Local Similarity   100.0%; Pred. No. 2.1e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
               ||||
Db   23 KAAK 26

RESULT 6
ID   Q9SPU2
AC   Q9SPU2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)

PRELIMINARY;   PRT;   35 AA.

DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DNA METHYLTRANSFERASE (FRAGMENT).
GN   MET1
CS   Arabidopsis thaliana (Mouse-ear cress)
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20044094; PubMed=10579493;
RA   Genger R.K., Kovac K.A., Dennis E.S., Peacock W.J., Finnegan E.J.;
RT   "Multiple DNA methyltransferase genes in Arabidopsis thaliana.";
RL   Plant Mol. Biol. 41:269-278(1999).
DR   EMBL; AF139372; AAF01675.1; -.
KW   Transferase; Methyltransferase.
NON_TER   35   35
SQ   SEQUENCE   35 AA;   3983 MW;   4FCDF571B6BEC4DB CRC64;

Query Match   100.0%; Score 18; DB 10; Length 35;
Best Local Similarity   100.0%; Pred. No. 2.7e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
               ||||
Db   7 KAAK 10

RESULT   7
ID   031573
AC   031573;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-NOV-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   YFHE PROTEIN.
GN   YFHE.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,

---

RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97101647; PubMed=8946165;
RA   Yamamoto H., Uchiyama S., Sekiguchi J.;
RT   "Cloning and sequencing of a 27.8-kb nucleotide sequence of the 79
RT   degrees-81 degrees region of the Bacillus subtilis genome containing
RT   the sspB locus.";
RL   DNA Res. 3:257-262(1996).
DR   EMBL; Z99108; CAB12679.1; -.
DR   EMBL; D85082; BAA24471.1; -.
SQ   SEQUENCE   36 AA;   4277 MW;   2EE69B192DDD62C0 CRC64;

Query Match   100.0%; Score 18; DB 2; Length 36;
Best Local Similarity   100.0%; Pred. No. 2.7e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
               ||||
Db   31 KAAK 34

RESULT   8
ID   085035
AC   085035;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 4.4 KDA PROTEIN.
OS   Mycoplasma hominis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2098;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PG21;
RX   MEDLINE=98353329; PubMed=9687401;
RA   Bebear C.M., Charron A., Bove J.M., Bebear C., Renaudin J.;
RT   "Cloning and nucleotide sequences of the topoisomerase IV parC and
RT   parE genes of Mycoplasma hominis.";
RL   Antimicrob. Agents Chemother. 42:2024-2031(1998).
DR   EMBL; AF036961; AAC33843.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   37 AA;   4369 MW;   B661BD716143A9B4 CRC64;

Query Match   100.0%; Score 18; DB 2; Length 37;
Best Local Similarity   100.0%; Pred. No. 2.8e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
               ||||
Db   12 KAAK 15

RESULT   9
ID   061191
AC   061191;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)

```
DE   HISTONE H3 (FRAGMENT).
GN   H3-1.
OS   Blepharisma undulans.
OC   Eukaryota; Alveolata; Ciliophora; Spirotrichea; Heterotrichida;
OC   Blepharisma.
OX   NCBI_TaxID=74723;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98154410; PubMed=9493359;
RA   Bernhard D., Schlegel M.;
RT   "Evolution of histone H4 and H3 genes in different ciliate lineages.";
RL   J. Mol. Evol. 46:344-354(1998).
DR   EMBL; AJ004689; CAA06033.1; -.
FT   NON_TER       37      37
SQ   SEQUENCE      37 AA;   3884 MW;  B7F723ECF3F9C6F2 CRC64;

Query Match             100.0%;  Score 18;  DB 5;  Length 37;
Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KAAK 4
              ||||
Db         25 KAAK 28

RESULT 10
061189
ID   061189;    PRELIMINARY;       PRT;       38 AA.
AC   061189;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HISTONE H3 (FRAGMENT).
GN   H3-3.
OS   Blepharisma undulans.
OC   Eukaryota; Alveolata; Ciliophora; Spirotrichea; Heterotrichida;
OC   Blepharisma.
OX   NCBI_TaxID=74723;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98154410; PubMed=9493359;
RA   Bernhard D., Schlegel M.;
RT   "Evolution of histone H4 and H3 genes in different ciliate lineages.";
RL   J. Mol. Evol. 46:344-354(1998).
DR   EMBL; AJ004691; CAA06043.1; -.
FT   NON_TER       38      38
SQ   SEQUENCE      38 AA;   3914 MW;  E86D5AFF9CEF93C6 CRC64;

Query Match             100.0%;  Score 18;  DB 5;  Length 38;
Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KAAK 4
              ||||
Db         25 KAAK 28

RESULT 11
061190
ID   061190;    PRELIMINARY;       PRT;       38 AA.
AC   061190;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HISTONE H3 (FRAGMENT).
GN   H3-2.
OS   Blepharisma undulans.
OC   Eukaryota; Alveolata; Ciliophora; Spirotrichea; Heterotrichida;
OC   Blepharisma.
OX   NCBI_TaxID=74723;
RN   [1]


RP   SEQUENCE FROM N.A.
RX   MEDLINE=98154410; PubMed=9493359;
RA   Bernhard D., Schlegel M.;
RT   "Evolution of histone H4 and H3 genes in different ciliate lineages.";
RL   J. Mol. Evol. 46:344-354(1998).
DR   EMBL; AJ004690; CAA06041.1; -.
FT   NON_TER       38      38
SQ   SEQUENCE      38 AA;   3916 MW;  E86D5AFB8CF7F2C6 CRC64;

Query Match             100.0%;  Score 18;  DB 5;  Length 38;
Best Local Similarity   100.0%;  Pred. No. 2.9e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KAAK 4
              ||||
Db         25 KAAK 28

RESULT 12
Q29047
ID   Q29047     PRELIMINARY;       PRT;       39 AA.
AC   Q29047;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   G PROTEIN ALPHA I SUBUNIT (FRAGMENT).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE=93179393; PubMed=8440688;
RA   Holtzman E.J., Kinane T.B., West K., Soper B.W., Karga H.,
RA   Ausiello D.A., Ercolani L.;
RT   "Transcriptional regulation of G-protein alpha i subunit genes in LLC-
RT   PK1 renal cells and characterization of the porcine G alpha i-3 gene
RT   promoter.";
RL   J. Biol. Chem. 268:3964-3975(1993).
DR   EMBL; L07317; AAA31049.1; -.
DR   HSSP; P10824; 1GG2.
DR   INTERPRO; IPR001019; -.
DR   PFAM; PF00503; G-alpha; 1.
KW   GTP-binding.
FT   NON_TER       39      39
SQ   SEQUENCE      39 AA;   4333 MW;  B029D6305AFDB028 CRC64;

Query Match             100.0%;  Score 18;  DB 6;  Length 39;
Best Local Similarity   100.0%;  Pred. No. 3e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KAAK 4
              ||||
Db         29 KAAK 32

RESULT 13
Q47512
ID   Q47512     PRELIMINARY;       PRI;       45 AA.
AC   Q47512;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   PECTIN LYASE (PNL) (FRAGMENT).
OS   Erwinia carotovora.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pectobacterium.
OX   NCBI_TaxID=554;
RN   [1]
RP   SEQUENCE FROM N.A.
```

RX   MEDLINE-90241256; PubMed=2185758;
RA   Nishida T., Suzuki T., Ito K., Kamio Y., Izaki K.;
RT   "Cloning and expression of pectin lyase gene from Erwinia carotovora
RT   in Escherichia coli.";
RL   Biochem. Biophys. Res. Commun. 158:801-808(1990).
DR   EMBL; M35271; AAA24855.1; -.
KW   Lyase.
FT   NON_TER   45   45
SQ   SEQUENCE   45 AA;   4483 MW;   4316F4C06CD5BOD7 CRC64;

Query Match   100.0%; Score 18; DB 2; Length 45;
Best Local Similarity   100.0%; Pred. No. 3.4e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
              ||||
Db   16 KAAK 19

RESULT 14
ID   Q92NU5   PRELIMINARY;   PRT;   54 AA.
AC   Q92NU5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   RUVB PROTEIN (FRAGMENT).
GN   RUVB.
OS   Clostridium histolyticum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1498;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-JCM 1403;
RA   Matsushita O.;
RT   "Clostridium histolyticum collagenase.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB014075; BAA34546.1; -.
FT   NON_TER   54   54
SQ   SEQUENCE   54 AA;   6312 MW;   492B3EA9A8933292 CRC64;

Query Match   100.0%; Score 18; DB 2; Length 54;
Best Local Similarity   100.0%; Pred. No. 4e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
              ||||
Db   44 KAAK 47

RESULT 15
ID   O97314   PRELIMINARY;   PRT;   54 AA.
AC   O97314;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PFC1025C PROTEIN.
GN   PFC1025C
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-3D7;
RA   Oliver K., Bowman S., Harris D., Lawson D., Quail M., Barrell B.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL034559; CAB39063.1; -.
SQ   SEQUENCE   54 AA;   6065 MW;   E24DBCEB3DE2492B CRC64;

Query Match   100.0%; Score 18; DB 5; Length 54;
Best Local Similarity   100.0%; Pred. No. 4e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAK 4
              ||||
Db   14 KAAK 17

Search completed: April 24, 2001, 16:40:30
Job time: 420 sec

```
                GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      April 24, 2001, 16:38:17 ; Search time 115.25 seconds
                                        (without alignments)
                                        2.480 Million cell updates/sec

Title:       US-09-340-736-4
Perfect score: 22
Sequence:    1 KAAAK 5

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

tched:       390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters:      390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_0401:*
          1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
          2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
          3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
          4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
          5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
          6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
          7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
          8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
          9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
         10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
         11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
         12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
         13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
         14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
         15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
         16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
         17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
         18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
         19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
         20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
         21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
         22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	22	100.0	7	22	B37101	Poly(Lys)-poly(Ala
2	22	100.0	9	11	R04181	Monomer of repeat
3	22	100.0	9	20	W95595	Peptide structure
4	22	100.0	11	11	R08089	Antifreeze segment
5	22	100.0	11	12	W35172	Lys/Ala diastereom
6	22	100.0	12	19	W82888	Antipathogenic pep
7	22	100.0	12	19	B17423	Antipathogenic pep
8	22	100.0	13	20	Y34065	Histone H1 isoform
9	22	100.0	13	21	Y57363	Human histone H1.5
10	22	100.0	15	20	Y34063	Histone H1 isoform
11	22	100.0	15	21	Y57361	Human histone H1.5
12	22	100.0	16	17	W07279	Amphiphilic antimi
13	22	100.0	18	14	R32756	Murine epsilon RMP
14	22	100.0	19	14	R32755	Oligopeptide based
15	22	100.0	19	21	Y87835	Heparin binding pe
16	22	100.0	21	16	R76397	Bacterial permeabi
17	22	100.0	23	17	W01634	N-dansylated pepti
18	22	100.0	25	21	Y91770	Amino acid sequenc
19	22	100.0	27	21	B18735	Amino acid sequenc
20	22	100.0	29	21	B18736	Amino acid sequenc
21	22	100.0	36	19	B18727	Amino acid sequenc
22	22	100.0	39	19	R44934	Mycobacterial hepa
23	22	100.0	45	20	Y62572	Copolymer molecula
24	22	100.0	55	20	Y34069	Histone H1 isoform
25	22	100.0	55	21	Y57367	Human histone H1.5
26	22	100.0	56	21	G35275	Zea mays protein f
27	22	100.0	59	21	Y82373	Copolymer molecula
28	22	100.0	59	21	G35274	Zea mays protein f
29	22	100.0	60	21	B21449	Cone snail alpha-c
30	22	100.0	61	19	W77601	Staphylococcus aur
31	22	100.0	63	21	G35273	Zea mays protein f
32	22	100.0	66	21	Y82574	Copolymer molecula
33	22	100.0	77	20	Y60367	Human normal bladd
34	22	100.0	77	21	Y82575	Copolymer molecula
35	22	100.0	86	21	Y82576	Arabidopsis thalia
36	22	100.0	91	21	G55059	Arabidopsis thalia
37	22	100.0	105	21	G03935	Human secreted pro
38	22	100.0	109	21	Y82577	Copolymer molecula
39	22	100.0	113	21	G27987	Arabidopsis thalia
40	22	100.0	113	21	G47670	Arabidopsis thalia
41	22	100.0	116	20	Y34066	Histone H1 isoform
42	22	100.0	116	21	Y57364	Human histone H1.5
43	22	100.0	119	21	Y52542	Murine H2-M/multip
44	22	100.0	124	22	G26549	Arabidopsis thalia
45	22	100.0	125	20	Y37409	Protein involved i

ALIGNMENTS

```
RESULT   1
B37101
ID   B37101 standard; peptide; 7 AA.
XX
AC   B37101;
XX
DT   28-MAR-2001  (first entry)
XX
DE   Poly(Lys)-poly(Ala) motif R2 for protein purification method.
XX
KW   Production; purification; hybrid polypeptide; affinity chromatography;
KW   sulphated polysaccharide; binding motif; antigen; vaccine.
XX
OS   Synthetic.
XX
PN   WO200068398-A1.
XX
PD   16-NOV-2000.
XX
PF   11-MAY-1999; 2000WO-FR01282.
XX
PR   11-MAY-1999;  99FR-0006031.
XX
PA   (INSP ) INST PASTEUR LILLE.
PA   (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI   Menozzi F, Locht C,  Pethe K;
XX
DR   WPI; 2001-007400/01.
XX
PT   Purifying hybrid proteins by affinity chromatography, useful for
PT   preparing vaccinating antigens, by incorporating motifs with affinity
PT   for sulfated polysaccharide -
XX
```

PS   Claim 4; Page 44; 6lpp; French.
XX
CC   The invention relates to a method of producing and purifying a hybrid
CC   polypeptide (I) by affinity chromatography on a support carrying
CC   sulphated polysaccharides (SP). (I) consists of a protein (II), at least
CC   one motif (R1) of formula X1X2Y1Y2PY3 and 0-5 motifs (R2) of formula
CC   X1X2Y1Y2Y3X3X4, where X1-X4 = K or R: and Y1-Y3 = any amino acid, but
CC   preferably A, L, I, or V. The total number of R1 and R2 is chosen
CC   according to the affinity required between (I) and SP, provided that if
CC   no R2 are present there must be at least two R1. Optionally purified (I)
CC   is treated with a serine protease to remove R1 and R2, leaving (II). This
CC   peptide sequence represents an example of an R2 motif used in the hybrid
CC   polypeptide. The method comprises inserting the coding sequence for (I)
CC   into a vector able to express (I). The vector is introduced into suitable
CC   host cells, then expressed (I) is contacted with immobilized SP and bound
CC   (I) eluted with a 0-1 M salt gradient. The method is particularly used
CC   where (I) includes a vaccinating antigen (Ag), optionally used with
CC   proteolytic removal of R1 and R2 for recovery of Ag.
XX
SQ   Sequence   7 AA;

Query Match   100.0%;   Score 22; DB 22; Length 7;
1st Local Similarity   100.0%;   Pred. No. 3.2e+05;
Matches   5;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 KAAAK 5
        |||||
Db   2 kaaak 6

RESULT 2
R04181
ID   R04181 standard; protein; 9 AA.
XX
AC   R04181;
XX
DT   10-SEP-1990 (first entry)
XX
DE   Monomer of repetitive primary structure of a synthetic fibrous
DE   protein.
XX
KW   Collagen; elastin; keratin; troponin C; silk; dopa protein;
KW   synthetic skin; cosmetics; bioadhesive.
XX
OS   Synthetic.
XX
PN   WO9003438-A.
XX
PD   05-APR-1990.
XX
PR   06-SEP-1989; 89WO-0003839.
PR   30-SEP-1988; 88US-0251714.
XX
PA   (ALLC) ALLIED SIGNAL INC.
XX
PI   Goldberg I, Salerno AJ;
XX
DR   WPI; 1990-132274/17.
XX
PT   New bacterial strains for heterologous gene expression -
PT   contg. elements for initiating activity, retarding proteolysis
PT   and stabilising heterologous genes.
XX
PS   Disclosure; ; 1pp; English.
XX
CC   The polymer of the peptide may act as an analogue to peptides
CC   contg. a repetetive or quasi-repetetive structure eg. collagen,
CC   elastin, keratin, troponin C, dopa proteins, silk proteins,
CC   bioadhesive proteins and insect cuticle proteins. These products
CC   may be useful in the manufacture of fibrous products, synthetic
CC   skin and cosmetic additives.

XX
SQ   Sequence   9 AA;

Query Match   100.0%;   Score 22; DB 11; Length 9;
Best Local Similarity   100.0%;   Pred. No. 3.2e+05;
Matches   5;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 KAAAK 5
        |||||
Db   3 kaaak 7

RESULT 3
W95595
ID   W95595 standard; peptide; 9 AA.
XX
AC   W95595;
XX
DT   31-MAR-1999 (first entry)
XX
DE   Peptide structure depicting a peptidoglycan (mucopeptide) complex.
XX
KW   Mucopeptide; MPI; lipopolysaccharide; LPS; inhibitory; septic shock;
KW   arthritis; Group A mucopeptide; inflammatory disease; rheumatoid;
KW   Crohn's disease; psoriasis; peptidoglycan.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Modified-site   1
FT   /note= "attached to a sugar complex"
FT   Modified-site   2
FT   /note= "D-form residue; Glu-NH2"
FT   Modified-site   3
FT   /note= "attached to D-Ala which is further connected
FT   to another D-Ala (both residues not shown in
FT   the sequence)"
FT   Misc-difference 6
FT   /note= "D-form residue"
FT   Misc-difference 8
FT   /note= "D-form residue"
FT   Modified-site   9
FT   /note= "Ala-(MurNac-GlcNac)"
XX
PN   WO9857657-A1.
XX
PD   23-DEC-1998.
XX
PF   17-JUN-1998; 98WO-US12647.
XX
PR   16-JUN-1998; 98US-0097838.
PR   18-JUN-1997; 97US-0050087.
PR   18-JUN-1997; 97US-0877988.
XX
PA   (UIRQ) UNIV ROCKEFELLER.
XX
PI   Blake MS, Zabriskie JB;
XX
DR   WPI; 1999-080932/07.
XX
PT   Identifying peptides with affinity to anti-Group A mucopeptide
PT   antibodies - useful to treat septic shock and experimental arthritis
XX
PS   Disclosure; Fig 3; 47pp; English.
XX
CC   The invention relates to a new method of identifying an inhibitory
CC   peptide. The method comprises contacting a selected peptide able to bind
CC   a Group A mucopeptide (MPI) and a lipopolysaccharide (LPS) molecule
CC   with: (a) MPI; and (b) a mucopeptide antibody (AB1) and selecting a
CC   peptide which inhibits binding of MPI with AB1. The inhibitory peptide
CC   obtained by the method is used to treat septic shock in mammals. AB1 is
CC   used to prevent or treat arthritis in animals. The selected peptide is

CC used in a method for identifying antigenic peptides that are used to
CC prevent or ameliorate experimental arthritis induced in an animal. In its
CC broadest application the antibodies raised against group A. mucopeptide
CC can be used for the treatment of inflammatory diseases such as septic
CC shock, and rheumatoid arthritis, Crohn's disease, psoriasis, and for the
CC identification of specific peptides that can be used in such treatment.
CC The present sequence represents the structure of a peptidoglycan
CC (mucopeptide) complex.
XX
SQ Sequence 9 AA;

Query Match 100.0%; Score 22; DB 20; Length 9;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAAK 5
     | | | | |
Db 3 kaaak 7

RESULT 4
ID R08089 standard; protein; 11 AA.
AC R08089;
XX
DT 01-MAR-1991 (first entry)
XX
DE Antifreeze segment #3 encoded by S8-1.
XX
KW synthetic antifreeze polypeptide; cryopreservation;
KW core repetitive sequence.
XX
OS Synthetic.
XX
PN WO9013571-A.
XX
PD 15-NOV-1990.
XX
PF 10-MAY-1990; 90WO-US02626.
XX
PR 10-MAY-1989; 89US-0350481.
PR 10-MAY-1990; 90US-0507716.
XX
PA (DNAP-) DNA PLANT TECHN COR.
XX
PI Warren GJ, Mueller GM, Mckown RL, Dunsmuir P;
XX
DR WPI; 1990-361428/48.
XX
PT New anti-freeze polypeptides - used for cryo-protection of e.g.
PT foods, medically used biologics, plant prods. or plants during
PT growth.
XX
PS Disclosure: Fig 4; 11pp; English.
XX
CC Synthetic anti-freeze polypeptides (saf) comprise a specifically
CC cleavable site and a region containing at least two "core" segments
CC such as the 11mer given here. The saf's suppress ice crystal growth
CC by binding to the growing crystal face and blocking sites for
CC further crystal growth. They can be used to maximise retention of
CC important properties of organic materials through freezing and
CC thawing processes. The basic design of the polypeptides is based
CC on known antifreeze polypeptides from insects and fish, e.g. the
CC winter flounder.
CC See also Q06649-Q06652, Q05656, Q06656, Q06658, Q06672, R08070-6,
CC R08087-8, R08090-R08093.
XX
SQ Sequence 11 AA;

Query Match 100.0%; Score 22; DB 11; Length 11;

Best Local Similarity 100.0%; Pred. No. 21; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAAK 5
     | | | | |
Db 7 kaaak 11

RESULT 5
ID W35172 standard; peptide; 12 AA.
AC W35172;
XX
DT 14-APR-1998 (first entry)
XX
DE Lys/Ala diastereomer peptide [D]-A3,4,8,10-K4A8.
XX
KW Lys/Ala diastereomer peptide; infection; therapy; excitatory neurotoxin;
KW Honey bee venom; pardaxin; cytolytic activity; cancer;
KW non-haemolytic; preservative; agricultural produce; bacterial cell lysis;
KW agricultural pesticide; cell wall lysis.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 3
FT /note= "D-form residue"
FT Misc-difference 4
FT /note= "D-form residue"
FT Misc-difference 8
FT /note= "D-form residue"
FT Misc-difference 10
FT /note= "D-form residue"
FT Modified-site 12
FT /note= "C-terminal amide"
XX
PN WO9731019-A2.
XX
PD 28-AUG-1997.
XX
PF 20-FEB-1997; 97WO-IL00066.
XX
PR 22-FEB-1996; 96IL-0117223.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
XX
PI Oren Z, Shai Y;
XX
DR WPI; 1997-435088/40.
XX
PT Peptide(s) having selective cytolytic activity - against pathogens
PT and malignant cells, but no haemolytic activity, used for treating
PT infections and cancer
XX
PS Example 4; Page 45; 80pp; English.
XX
CC This sequence represents a Lys/Ala diastereomer peptide of the
CC invention. The peptides of the invention have: (a) cytolytic activity on
CC pathogenic cells (pathogens and malignant cells not naturally present in
CC the body); but (b) no haemolytic activity, or such activity only at a
CC concentration significantly higher than that at which they lyse
CC pathogens. The peptides, their complexes and mixtures are used to treat
CC infections (caused by bacteria, fungi, protozoa, mycoplasma or viruses)
CC or cancer, in human and veterinary medicine. Also, they can be used as
CC preservatives for food, cosmetics and agricultural produce, or as
CC agricultural pesticides. The absence of haemolytic activity (associated
CC with disturbance of alpha-helical structures) means that the peptides
CC have few if any toxic effects, and those that include D-aa will have
CC increased resistance to proteolytic degradation. Non-haemolytic.
CC cytotoxic random copolymers of pardaxin, each has a specific spectrum of
CC activity, allowing selection of agents for particular applications. Since
CC these random copolymers induce total lysis of bacterial cell walls,

CC resistance to them is unlikely to develop.
XX
SQ Sequence 12 AA;

Query Match   100.0%; Score 22; DB 18; Length 12;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAaK 5
      |||||
Db 1 kaaak 5

RESULT 6
W82888
ID W82888 standard; peptide; 12 AA.
XX
AC W82888;
XX
DT 19-MAY-1999 (first entry)
XX
DE Antipathogenic peptide.
XX
KW Non-haemolytic; cytolytic; selective cytolytic activity; pathogen;
KW cancer; infection; disinfectant; contact lens wetting solution;
KW preservative; pesticide; fungicide; bactericide.
XX
OS Synthetic.
XX
PN WO9837090-A1.
XX
PD 27-AUG-1998.
XX
PF 19-FEB-1998; 98WO-IL00081.
XX
PR 20-FEB-1997; 97WO-IL00066.
XX
PA (YEDA) YEDA RES & DEV CO LTD.
XX
PI Oren Z, Shai Y;
XX
DR WPI; 1998-594464/50.
XX
PT New non-haemolytic cytolytic agent useful in treating cancer or
PT infections - is a peptide comprising a moiety which disrupts the
PT continuity of an alpha-helical structure
XX
PS Example 4; Page 44; 126pp; English.
XX

CC The present peptide is used to produce the agents of the invention. The
CC specification describes a non-haemolytic, cytolytic agent, which is a
CC peptide, a complex of bundled peptides, a mixture of peptides or a random
CC peptide copolymer. The agent has a selective cytolytic activity on
CC pathogenic cells. The agent is selected from a cyclic derivative of a
CC peptide which has a net positive charge greater than 1, comprises L-amino
CC acid residues and/or D-amino acid residues and comprises an alpha-helix
CC breaker moiety, or a peptide (or cyclic derivative of this) which
CC (comprises L-amino acid residues and D-amino acid residues, has a net
CC positive charge greater than 1 and has an amino acid sequence such that
CC a corresponding amino acid sequence comprising only L-amino acid residues
CC is not found in nature. The cytolytic agents may be used for treatment of
CC cancer or for treatment of several diseases caused by pathogens,
CC including bacterial, fungal, viral, mycoplasma and protozoan infections.
CC They may be used in both human and veterinary medicine. They may also be
CC used as disinfectants for destruction of microorganisms, i.e. in
CC solutions for wetting contact lenses, as preservatives, e.g. in the
CC cosmetic and food industries, as pesticides (e.g. fungicides or
CC bactericides) or for preservation of agricultural products.
XX
SQ Sequence 12 AA;

Query Match   100.0%; Score 22; DB 19; Length 12;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAaK 5
      |||||
Db 1 kaaak 5

RESULT 7
E17423
ID E17423 standard; Peptide; 12 AA.
XX
AC E17423;
XX
DT 31-OCT-2000 (first entry)
XX
DE Antipathogenic peptide sequence SEQ ID NO:527.
XX
KW Modified peptide; therapeutic agent; fusion; Fc domain; cancer;
KW autoimmune disease; disease; cytostatic; antiasthmatic; thrombolytic; VEGF;
KW immunosuppressive; EPO; TPO; CTLA4; mimetic; IL-1; TNF; antagonist;
KW MMP; inhibitor; erythropoietin; thrombopoietin; interleukin 1;
KW cytotoxic T cell lymphocyte antigen 4; tumour necrosis factor;
KW vascular endothelial growth factor; matrix metalloproteinase;
KW asthma; thrombosis; pharmaceutical.
XX
OS Synthetic.
XX
PN WO200024782-A2.
XX
PD 04-MAY-2000.
XX
PF 25-OCT-1999; 99WO-US25044.
XX
PR 23-OCT-1998; 98US-0105371.
PR 22-OCT-1999; 99US-0428082.
XX
PA (AMGE-) AMGEN INC.
XX
PI Feige U, Liu C, Cheetham J, Boone TC;
XX
DR WPI; 2000-350702/30.
XX
PT Novel composition of matter comprising an Fc domain and
PT pharmacologically active peptides, useful for treating cancer and
PT autoimmune diseases -
XX
PS Claim 39; Page 381; 608pp; English.
XX

CC The present invention describes composition of matter (I) comprising an
CC Fc domain, pharmacologically active peptides, and linkers. Where (I) is:
CC (X1)a-F1-(X2)b, where: F1 = an Fc domain; X1 and X2 are each
CC independently selected from -(L1)c-P1, -(L1)c-P1-(L2)d-P2,
CC -(L1)c-P1-(L2)d-P2-(L3)e-P3, or -(L1)c-P1-(L2)d-P2-(L3)e-P3-(L4)f-P4
CC where P1, P2, P3, and P4 are each independently sequences of
CC pharmacologically active peptides; L1, L2, L3, and L4 = are each
CC independently linkers; and a, b, c, d, e, and f = are each independently
CC 0 or 1, provided that at least 1 of a and b is 1. The composition can
CC have cytostatic, antiasthmatic, thrombolytic and immunosuppressive
CC activities. DNAs, vectors and host cells from the present invention can
CC be used for producing pharmaceutical compositions. The compositions are
CC useful for treating cancer, asthma, thrombosis, or autoimmune diseases.
CC The use of an Fc domain (rather than a Fab domain) can provide a longer
CC half-life or incorporate functions such as Fc receptor binding, protein A
CC binding, complement fixation, and possibly placental transfer. A69443 to
CC A69526 and E16955 to E18003 represent nucleotide and amino acid sequences
CC used in the exemplification of the present invention.
XX
SQ Sequence 12 AA;

Query Match   100.0%; Score 22; DB 21; Length 12;

Best Local Similarity 100.0%; Pred. No. 23;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 KAAAK 5
        |||||
Db    1 Kaaak 5
```

RESULT 8
Y34065
ID Y34065 standard; peptide; 13 AA.
XX
AC Y34065;
XX
DT 23-NOV-1999 (first entry)
XX
DE Histone H1 isoform H1.5 pANCA-reactive fragment (residues 172-184).
XX
KW Ulcerative colitis; histone; H1-like antigen; porin antigen; human;
KW Bacteroides antigen; inflammatory bowel disease; IBD; pANCA; diagnosis;
KW perinuclear anti-neutrophil cytoplasmic antibody; isoform.
XX
OS Homo sapiens.
XX
PN WO9945955-A1.
XX
PD 16-SEP-1999.
XX
PF 12-MAR-1999; 99WO-US05492.
XX
PR 12-MAR-1998; 98US-0041889.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Braun J, Cohavy O;
XX
DR WPI; 1999-551215/46.
XX
PT Use of histone H1, porin or Bacteroides antigens as targets for the
PT diagnosis, prevention and treatment of ulcerative colitis -
XX
PS Example 4; Page 124; 134pp; English.
XX
CC The invention provides a method for the diagnosis, prevention and
CC treatment of ulcerative colitis (UC) using histone H1-like antigen, a
CC porin antigen or a Bacteroides antigen as a target antigen. The novel
CC method of diagnosing UC in a subject suspected of having inflammatory
CC bowel disease (IBD) comprises: (1) obtaining a sample from the subject;
CC (2) contacting the sample with a histone H1-like antigen, or perinuclear
CC anti-neutrophil cytoplasmic antibody (pANCA) -reactive fragment, to form a
CC complex of the histone H1-like antigen, or the pANCA-reactive fragment,
CC and antibody to the histone H1-like antigen; and (3) detecting the
CC presence or absence of the complex; where the presence of the complex
CC indicates that the subject has UC. The pANCA-reactive histone H1-like
CC antigen, porin antigen and Bacteroides antigen are useful in the
CC diagnosis, prevention and treatment of UC. The methods can also be used
CC for identifying agents useful for treating UC. The present sequence
CC represents a human histone H1 isoform H1.5 pANCA-reactive fragment.
XX
SQ Sequence 13 AA;

Query Match   100.0%; Score 22; DB 20; Length 13;
Best Local Similarity 100.0%; Pred. No. 25;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 KAAAK 5
        |||||
Db    7 kaaak 11
```

RESULT 9
Y57363

ID Y57363 standard; peptide; 13 AA.
XX
AC Y57363;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human histone H1.5 pANCA-reactive fragment (residues 172-184).
XX
KW Ulcerative colitis; inflammatory bowel disease; porin antigen;
KW pANCA; perinuclear anti-neutrophil cytoplasmic antibody; human;
KW histone H1; isoform; NANUC-2.
XX
OS Homo sapiens.
XX
PN US6033864-A.
XX
PD 07-MAR-2000.
XX
PF 12-MAR-1998; 98US-0041889.
XX
PR 12-APR-1996; 96US-0057846.
PR 11-APR-1997; 97US-0037058.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Cohavy O, Braun J;
XX
DR WPI; 2000-255695/22.
XX
PT Diagnosing ulcerative colitis or susceptibility, by detecting complex
PT formation between microbial porin antigen and perinuclear
PT anti-neutrophil cytoplasmic autoantibodies -
XX
PS Example 4; Columns 67-68; 49pp; English.
XX
CC The invention provides a method for diagnosing ulcerative colitis in a
CC subject suspected of having inflammatory bowel disease. The method
CC comprises reacting a patient sample with a porin antigen that is
CC immunologically reactive with pANCA (perinuclear anti-neutrophil
CC cytoplasmic antibodies) and detecting formation of a Ag-pANCA complex
CC as indicative of ulcerative colitis. The method is used to diagnose
CC ulcerative colitis or susceptibility to it. Sequences Y57362-367
CC represent pANCA-reactive fragments, derived from human histone H1.5.
XX
SQ Sequence 13 AA;

Query Match   100.0%; Score 22; DB 21; Length 13;
Best Local Similarity 100.0%; Pred. No. 25;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 KAAAK 5
        |||||
Db    7 kaaak 11
```

RESULT 10
Y34063
ID Y34063 standard; peptide; 15 AA.
XX
AC Y34063;
XX
DT 23-NOV-1999 (first entry)
XX
DE Histone H1 isoform H1.5 pANCA-reactive fragment (residues 170-184).
XX
KW Ulcerative colitis; histone; H1-like antigen; porin antigen; human;
KW Bacteroides antigen; inflammatory bowel disease; IBD; pANCA; diagnosis;
KW perinuclear anti-neutrophil cytoplasmic antibody; isoform.
XX
OS Homo sapiens.
XX
PN WO9945955-A1.

```
XX
PD   16-SEP-1999.
XX
PF   12-MAR-1999;   99WO-US05492.
XX
PR   12-MAR-1998;   98US-0041889.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Braun J, Cohavy O;
XX
DR   WPI; 1999-551215/46.
XX
PT   Use of histone H1, porin or Bacteroides antigens as targets for the
PT   diagnosis, prevention and treatment of ulcerative colitis   -
XX
PS   Example 4; Page 123; 134pp; English.
XX
CC   The invention provides a method for the diagnosis, prevention and
CC   treatment of ulcerative colitis (UC) using histone H1-like antigen, a
CC   porin antigen or a Bacteroides antigen as a target antigen. The novel
CC   method of diagnosing UC in a subject suspected of having inflammatory
CC   bowel disease (IBD) comprises: (1) obtaining a sample from the subject;
CC   (2) contacting the sample with a histone H1-like antigen, or perinuclear
CC   anti-neutrophil cytoplasmic antibody (pANCA)-reactive fragment, to form a
CC   complex of the histone H1-like antigen, or the pANCA-reactive fragment,
CC   and antibody to the histone H1-like antigen; and (3) detecting the
CC   presence or absence of the complex; where the presence of the complex
CC   indicates that the subject has UC. The pANCA-reactive histone H1-like
CC   antigen, porin antigen and Bacteroides antigen are useful in the
CC   diagnosis, prevention and treatment of UC. The methods can also be used
CC   for identifying agents useful for treating UC. The present sequence
CC   represents a human histone H1 isoform H1.5 pANCA-reactive fragment.
XX
SQ   Sequence   15 AA;

Query Match             100.0%;   Score 22;   DB 20;   Length 15;
Best Local Similarity   100.0%;   Pred. No. 28;
Matches   5;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     1 KAAAK 5
         |||||
Db     9 kaaak 13

RESULT 11
Y57361
ID   Y57361 standard; peptide; 15 AA.
XX
AC   Y57361;
XX
DT   13-JUN-2000   (first entry)
XX
DE   Human histone H1.5 pANCA-reactive fragment (residues 170-184).
XX
KW   Ulcerative colitis; inflammatory bowel disease; porin antigen;
KW   pANCA; perinuclear anti-neutrophil cytoplasmic antibody; human;
KW   histone H1; isoform; NANUC-2.
XX
OS   Homo sapiens.
XX
PN   US6033864-A.
XX
PD   07-MAR-2000.
XX
PF   12-MAR-1998;   98US-0041889.
XX
PR   12-APR-1996;   96US-0057846.
PR   11-APR-1997;   97US-0837058.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
```

```
PI   Cohavy O,  Braun J;
XX
DR   WPI; 2000-255695/22.
XX
PT   Diagnosing ulcerative colitis or susceptibility, by detecting complex
PT   formation between microbial porin antigen and perinuclear
PT   anti-neutrophil cytoplasmic autoantibodies   -
XX
PS   Disclosure; Columns 55-66; 49pp; English.
XX
CC   The invention provides a method for diagnosing ulcerative colitis in a
CC   subject suspected of having inflammatory bowel disease. The method
CC   comprises reacting a patient sample with a porin antigen that is
CC   immunologically reactive with pANCA (perinuclear anti-neutrophil
CC   cytoplasmic antibodies) and detecting formation of a Ag-pANCA complex
CC   as indicative of ulcerative colitis. The method is used to diagnose
CC   ulcerative colitis or susceptibility to it. Sequences Y57362-367
CC   represent pANCA-reactive fragments, derived from human histone H1.5.
XX
SQ   Sequence   15 AA;

Query Match             100.0%;   Score 22;   DB 21;   Length 15;
Best Local Similarity   100.0%;   Pred. No. 28;
Matches   5;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     1 KAAAK 5
         |||||
Db     9 kaaak 13

RESULT 12
W07279
ID   W07279 standard; peptide; 16 AA.
XX
AC   W07279;
XX
DT   29-MAY-1997   (first entry)
XX
DE   Amphiphilic antimicrobial peptide MB-34.
XX
KW   amphiphilic antimicrobial; log kill; Staphylococcus aureus; charge;
KW   Escherichia coli; vectorial analysis; hydrophile; lipophile balance;
KW   hydrophobic moment; equation; antimicrobial; antibacterial;
KW   antifungal; disinfection; spoilage prevention; preservation.
XX
OS   Synthetic.
XX
PN   WO9628468-A2.
XX
PD   19-SEP-1996.
XX
PF   27-FEB-1996;   96WO-EP00844.
XX
PR   09-MAR-1995;   95GB-0004761.
XX
PA   (UNIL ) UNILEVER NV.
PA   (UNIL ) UNILEVER PLC.
XX
PI   Bhakoo M,  Patel S,  Stott PI;
XX
DR   WPI; 1996-433760/43.
XX
PT   New amphiphilic antimicrobial peptide(s) - having particular
PT   combinations of hydrophilic and hydrophobic amino acid residues
XX
PS   Example -; Page 22; 39pp; English.
XX
CC   WO7270-90 examples of amphiphilic antimicrobial peptides for comparison
CC   to claimed amphiphilic antimicrobial peptides (WO7257-69), and show that
CC   relatively minor changes in structure can have significant consequences
CC   as regards the antimicrobial activity of the molecules. The log kills
CC   were predicted against S. aureus and against E. coli, and values of more
```

CC   than 5 and 4 respectively, correspond to greater antimicrobial activity.
CC   Effective antimicrobial peptides are discriminated from ineffective
CC   peptides by a vectorial analysis on dimensions corresponding to charge,
CC   hydrophile/lipophile balance, hydrophobic moment and amphiphilicity of
CC   the peptides, where effective peptides fall into the region which is
CC   anti-clockwise of the amphiphilicity dimension and clockwise of the
CC   charge dimension. Effective peptides are further discriminated from
CC   ineffective peptides by means of an equation relating certain physical
CC   properties of the peptides to their biological activity against specific
CC   microorganisms. The peptides can be used in antimicrobial, pref.
CC   antibacterial or antifungal, compsns. They can be used for the
CC   disinfection of surfaces, spoilage prevention, preservation or other
CC   hygiene processes.
XX
SQ   Sequence 16 AA;

Query Match   100.0%; Score 22; DB 17; Length 16;
Best Local Similarity 100.0%; Pred. No. 30;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
         |||||
Db   2 kaaaK 6

RESULT 13
R32756
ID   R32756 standard; peptide; 18 AA.
XX
AC   R32756;
XX
DT   21-JUN-1993 (first entry)
XX
DE   Murine epsilon RMP internal sequence.
XX
KW   Epsilon receptor modulating protein; IgE; CD23; class switching;
KW   allergy inflammation; partial sequence.
XX
OS   Mus musculus.
XX
PN   WO9302696-A.
XX
PD   18-FEB-1993.
XX
PF   06-AUG-1992; 92WO-US06553.
XX
PR   07-AUG-1991; 91US-0741671.
XX
PA   (MEDI-) MEDICAL BIOLOGY INST.
XX
PI   Katz DH, Matsushita S;
XX
DR   WPI; 1993-076175/09.
XX
PT   Immunologically active liq. for e.g. research into allergy etc. -
PT   contg. epsilon-receptor modifying protein of specified
PT   characteristics, used opt. with specified
PT   immuno-globin-E-suppressive protein etc.
XX
PS   Disclosure; Page 34; 75pp; English.
XX
CC   The peptide represents an internal sequence of the 17 kD epsilon
CC   receptor modulating protein (RMP). The protein may be used to
CC   modify the immune response of a mammal. Epsilon RMP together with
CC   IL-4 can initiate class switching to IgE synthesis. The mols. are
CC   also useful for treating allergic reactions, e.g. inflammation and
CC   in immunological research and in immunological therapy on humans and
CC   other animals. See also R32755.
XX
SQ   Sequence 18 AA;

Query Match   100.0%; Score 22; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 34;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
         |||||
Db   13 kaaaK 17

RESULT 14
R32755
ID   R32755 standard; peptide; 19 AA.
XX
AC   R32755;
XX
DT   21-JUN-1993 (first entry)
XX
DE   Oligopeptide based on internal sequence of murine epsilon RMP.
XX
KW   Epsilon receptor modulating protein; IgE; CD23; class switching;
KW   allergy inflammation; partial sequence.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Modified-site   19
FT   /note= "amidated"
XX
PN   WO9302696-A.
XX
PD   18-FEB-1993.
XX
PF   06-AUG-1992; 92WO-US06553.
XX
PR   07-AUG-1991; 91US-0741671.
XX
PA   (MEDI-) MEDICAL BIOLOGY INST.
XX
PI   Katz DH, Matsushita S;
XX
DR   WPI; 1993-076175/09.
XX
PT   Immunologically active liq. for e.g. research into allergy etc. -
PT   contg. epsilon-receptor modifying protein of specified
PT   characteristics, used opt. with specified
PT   immuno-globin-E-suppressive protein etc.
XX
PS   Disclosure; Page 13; 75pp; English.
XX
CC   The synthetic sequence is a 19 amino acid sequence with cysteine
CC   inserted into its N-terminus whose sequence was based on the
CC   internal amino acid sequence of purified epsilon receptor modulating
CC   protein. The synthetic oligopeptide was conjugated to keyhole
CC   limpet haemocyanin and used to immunise 8 week old female Lewis rats.
CC   The spleens were fused with non secreting mouse B cell myeloma, PO.
CC   A B cell hybridoma clone secreted antibody that reacted with the
CC   immunogen and with native epsilon RMP.
CC   See also R32756.
XX
SQ   Sequence 19 AA;

Query Match   100.0%; Score 22; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 36;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
         |||||
Db   14 kaaaK 18

RESULT 15
Y87835

ID   I87835 standard; peptide; 19 AA.
XX
AC   Y87835;
XX
DT   01-SEP-2000 (first entry)
XX
DE   Heparin binding peptide Lys helix #1.
XX
KW   Heparin binding peptide; antagonist; cardiovascular; coagulant;
KW   bleeding wound; vascular anastomoses; leaking prosthetic vascular graft;
KW   protamine substitute; treatment.
XX
OS   Synthetic.
XX
PN   EP999219-A2.
XX
PD   10-MAY-2000.
XX
PF   01-OCT-1999;   99EP-0119514.
XX
PR   06-OCT-1998;   98US-0166930.
XX
PA   (COMM-) COMMONWEALTH BIOTECHNOLOGIES INC.
XX
PI   Harris RB,   Sobel M;
XX
DR   WPI; 2000-306006/27.
XX
PT   New heparin binding molecules, useful for reducing heparin content in a
PT   mammal by reducing the anticoagulant effects of heparin -
XX
PS   Example 1; Page 8; 39pp; English.
XX
CC   This invention describes novel heparin binding molecules (I). The
CC   molecules (I) are useful as heparin antagonist drugs for cardiovascular
CC   application and specifically neutralize heparin's conventional
CC   anticoagulant properties. (I) are also useful for counteracting actions
CC   of heparin locally e.g. in bleeding wounds, vascular anastomoses or
CC   leaking prosthetic vascular grafts. (I) is also useful combined in a
CC   pharmaceutical composition with insulin, as a substitute for protamine
CC   for use in treating diabetics. The heparin binding molecules (I)
CC   specifically neutralize heparin's conventional anticoagulant properties
CC   without causing deleterious hemodynamic side-effects or exacerbation of
CC   the proliferative vascular response to injury. (I) are short-duration,
CC   intravenous drugs to be used in elective or emergency situations which
CC   can safely and specifically neutralize heparin's proliferative response
CC   to injury. This sequence represents a heparin binding peptide described
CC   in the method of the invention.
XX
SQ   Sequence   19 AA;

Query Match   100.0%; Score 22; DB 21; Length 19;
Best Local Similarity   100.0%;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
          |||||
Db   15 kaaak 19

Search completed: April 24, 2001, 16:38:18
Job time: 418 sec

```
            GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   April 24, 2001, 16:36:23 ; Search time 52.39 Seconds
                                       (without alignments)
                                       1.540 Million cell updates/sec

Title:          US-09-340-736-4
Perfect score:  22
Sequence:       1 KAAAK 5

Scoring table:  BLOSUM62
                Gapop 10.0 . Gapext 0.5

ched:   185757 segs, 19210857 residues
```

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	22	100.0	5	2	US-08-911-364-4	Sequence 4, Appl
2	22	100.0	13	3	US-09-041-889-37	Sequence 37, Appl
3	22	100.0	15	3	US-09-041-889-35	Sequence 35, Appl
4	22	100.0	18	5	PCT-US92-06553-2	Sequence 2, Appl
5	22	100.0	19	2	US-08-660-592-5	Sequence 5, Appl
6	22	100.0	19	4	US-09-166-930A-5	Sequence 5, Appl
7	22	100.0	19	5	PCT-US92-06553-1	Sequence 1, Appl
8	22	100.0	21	1	US-08-473-244-57	Sequence 57, Appl
9	22	100.0	21	1	US-08-473-244-58	Sequence 58, Appl
10	22	100.0	21	1	US-08-473-244-59	Sequence 59, Appl
11	22	100.0	23	5	PCT-US96-06224-1	Sequence 1, Appl
12	22	100.0	55	1	US-09-041-889-41	Sequence 41, Appl
13	22	100.0	60	1	US-08-346-849-16	Sequence 16, Appl
14	22	100.0	60	2	US-08-293-284A-16	Sequence 16, Appl
15	22	100.0	116	1	US-09-041-889-38	Sequence 38, Appl
16	22	100.0	158	1	US-09-041-889-40	Sequence 40, Appl
17	22	100.0	174	1	US-08-911-853-19	Sequence 19, Appl
18	22	100.0	199	2	US-08-800-264A-11	Sequence 11, Appl
19	22	100.0	199	2	US-09-018-628-11	Sequence 11, Appl
20	22	100.0	199	3	US-09-273-378-11	Sequence 11, Appl
21	22	100.0	199	3	US-09-209-605-11	Sequence 11, Appl
22	22	100.0	201	2	US-08-911-354-2	Sequence 2, Appl
23	22	100.0	214	3	US-09-041-889-27	Sequence 27, Appl
24	22	100.0	218	3	US-08-837-058-4	Sequence 4, Appl
25	22	100.0	218	2	US-09-041-889-4	Sequence 54, Appl
26	22	100.0	219	2	US-08-557-309B-54	Sequence 54, Appl
27	22	100.0	222	3	US-09-041-889-3	Sequence 3, Appl
28	22	100.0	222	3	US-08-837-058-3	Sequence 3, Appli
29	22	100.0	223	4	US-09-095-855-201	Sequence 201, App
30	22	100.0	226	2	US-09-041-889-32	Sequence 32, Appl
31	22	100.0	262	1	US-08-403-379A-1	Sequence 1, Appli
32	22	100.0	262	2	US-08-929-414-1	Sequence 1, Appli
33	22	100.0	263	2	US-08-557-309B-51	Sequence 51, Appl
34	22	100.0	263	2	US-08-834-306-51	Sequence 51, Appl
35	22	100.0	310	1	US-08-363-255-6	Sequence 6, Appli
36	22	100.0	320	2	US-08-800-264A-6	Sequence 6, Appli
37	22	100.0	320	2	US-09-018-628-6	Sequence 6, Appli
38	22	100.0	320	3	US-09-273-378-6	Sequence 6, Appli
39	22	100.0	320	3	US-09-209-605-6	Sequence 6, Appli
40	22	100.0	325	2	US-08-800-264A-8	Sequence 8, Appli
41	22	100.0	325	3	US-09-018-628-8	Sequence 8, Appli
42	22	100.0	325	3	US-09-273-378-8	Sequence 8, Appli
43	22	100.0	325	3	US-09-209-605-8	Sequence 8, Appli
44	22	100.0	366	2	US-08-804-699-2	Sequence 2, Appli
45	22	100.0	437	2	US-08-800-264A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-911-364-4
; Sequence 4, Application US/08911364
; Patent No. 5969106
; GENERAL INFORMATION:
; APPLICANT: ROTHSTEIN, Aser
; APPLICANT: KEELY, Fred W.
; APPLICANT: ROTHSTEIN, Steven J.
; TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
; TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,364
; FILING DATE: 07-AUG-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,552
; FILING DATE: 07-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 041082/0104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-911-364-4

Query Match   100.0%; Score 22; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 KAAAK 5
        |||||
Db    1 KAAAK 5

RESULT 2
US-09-041-889-37
; Sequence 37, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
;   APPLICANT: Braun, Jonathan
;   APPLICANT: Cohavy, Offer
;   TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
;   TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
;   TITLE OF INVENTION: Microbial UC pANCA antigens
;   NUMBER OF SEQUENCES: 41
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Campbell & Flores LLP
;     STREET: 4370 La Jolla Village Drive, Suite 700
;     CITY: San Diego
;     STATE: California
;     COUNTRY: USA
;     ZIP: 92122
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/041,889
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/837,058
;     FILING DATE: 11-APR-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Campbell, Cathryn A.
;     REGISTRATION NUMBER: 31,815
;     REFERENCE/DOCKET NUMBER: P-PM 3006
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (619) 535-9001
;     TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 37:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 13 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-09-041-889-37

Query Match             100.0%;  Score 22;  DB 3;  Length 13;
Best Local Similarity   100.0%;  Pred. No. 16;
Matches    5;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 KAAAK 5
        |||||
Db    7 KAAAK 11

RESULT 3
US-09-041-889-35
; Sequence 35, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
;   APPLICANT: Braun, Jonathan
;   APPLICANT: Cohavy, Offer
;   TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
;   TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
;   TITLE OF INVENTION: Microbial UC pANCA antigens
;   NUMBER OF SEQUENCES: 41
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Campbell & Flores LLP
;     STREET: 4370 La Jolla Village Drive, Suite 700
;     CITY: San Diego
;     STATE: California
;     COUNTRY: USA
;     ZIP: 92122
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/041,889
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/837,058
;     FILING DATE: 11-APR-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Campbell, Cathryn A.
;     REGISTRATION NUMBER: 31,815
;     REFERENCE/DOCKET NUMBER: P-PM 3006
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (619) 535-9001
;     TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 35:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 15 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-09-041-889-35

Query Match             100.0%;  Score 22;  DB 3;  Length 15;
Best Local Similarity   100.0%;  Pred. No. 18;
Matches    5;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 KAAAK 5
        |||||
Db    9 KAAAK 13

RESULT 4
PCT-US92-06553-2
; Sequence 2, Application PC/TUS9205553
; GENERAL INFORMATION:
;   APPLICANT: Katz M.D., David H
;   APPLICANT: Matsushita M.D., Sho
;   TITLE OF INVENTION: T-Cell Membrane Protein for B Cell IG
;   TITLE OF INVENTION: Class Switching
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Grant L. Hubbard
;     STREET: 2326 North Seventh Street
;     CITY: Phoenix
;     STATE: AZ
;     COUNTRY: USA
;     ZIP: 85006
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US92/06553
;     FILING DATE: 19520806
;     CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/741,671
;     FILING DATE: 07-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Hubbard, Grant L
```

; REGISTRATION NUMBER: 24193
; REFERENCE/DOCKET NUMBER: MBI032PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (602) 263-8782
; TELEFAX: (602) 265-5984
INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; CELL TYPE: T-cell
; CELL LINE: MBI-1.15
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: US 07/741671
; FILING DATE: 07-AUG-1991
US92-06553-2

Query Match   100.0%; Score 22; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 22; Mismatches 0;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
       |||||
Db 13 KAAAK 17

RESULT 5
US-08-660-592-5
; Sequence 5, Application US/08660592
; Patent No. 5877153
; GENERAL INFORMATION:
; APPLICANT: HARRIS, Robert B.
; APPLICANT: SOBEL, Michael
; TITLE OF INVENTION: NOVEL HEPARIN BINDING PEPTIDES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,592
; FILING DATE: 11-JUN-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 006338-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-660-592-5

Query Match   100.0%; Score 22; DB 2; Length 19;

Best Local Similarity 100.0%; Pred. No. 23;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
       |||||
Db 15 KAAAK 19

RESULT 6
US-09-166-930A-5
; Sequence 5, Application US/09166930A
; Patent No. 6200955
; GENERAL INFORMATION:
; APPLICANT: HARRIS, Robert B.
; APPLICANT: SOBEL, Michael
; TITLE OF INVENTION: NOVEL HEPARIN BINDING PEPTIDES
; FILE REFERENCE: 006338-006
; CURRENT APPLICATION NUMBER: US/09/166,930A
; CURRENT FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: US 08/660,592
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 5:
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:branched-chain
; OTHER INFORMATION: heparin-binding peptide Lys Helix #1
US-09-166-930A-5

Query Match   100.0%; Score 22; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 23;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
       |||||
Db 15 KAAAK 19

RESULT 7
PCT-US92-06553-1
; Sequence 1, Application PC/TUS9206553
; GENERAL INFORMATION:
; APPLICANT: Katz M.D., David H
; APPLICANT: Matsushita M.D., Sho
; TITLE OF INVENTION: T-Cell Membrane Protein for B Cell IG
; TITLE OF INVENTION: Class Switching
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grant L. Hubbard
; STREET: 2326 North Seventh Street
; CITY: Phoenix
; STATE: AZ
; COUNTRY: USA
; ZIP: 85006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06553
; FILING DATE: 19920806
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,671
; FILING DATE: 07-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hubbard, Grant L
; REGISTRATION NUMBER: 24193

REFERENCE/DOCKET NUMBER: MBI032PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (602) 263-8782
TELEFAX: (602) 265-5984
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: Peptide
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
CELL TYPE: T-helper cell, hybridoma
CELL LINE: MBI - 1.15
PUBLICATION INFORMATION:
DOCUMENT NUMBER: US 07/741671
FILING DATE: 07-AUG-1991
PCT-US92-06553-1

Query Match 100.0%; Score 22; DB 5; Length 19;
t Local Similarity 100.0%; Pred. No. 23;
tches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAAK 5
     |||||
Db 14 KAAAK 18

RESULT 8
US-08-473-344-57
; Sequence 57, Application US/08473344
; Patent No. 5763567
GENERAL INFORMATION:
APPLICANT: Little, Roger G
TITLE OF INVENTION: Biologically Active Peptides from
TITLE OF INVENTION: Functional Domains of Bactericidal/Permeability-Increasing
TITLE OF INVENTION: Protein and Uses Thereof
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,344
FILING DATE: 7-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/306,473
FILING DATE: 15-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,762
FILING DATE: 11-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: McDonnell, John J.
REGISTRATION NUMBER: 26,949
REFERENCE/DOCKET NUMBER: 93,1133-J
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 910-221-5317
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: "BPI 46"
US-08-473-344-57

Query Match 100.0%; Score 22; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAAK 5
     |||||
Db 3 KAAAK 7

RESULT 9
US-08-473-344-58
; Sequence 58, Application US/08473344
; Patent No. 5763567
GENERAL INFORMATION:
APPLICANT: Little, Roger G
TITLE OF INVENTION: Biologically Active Peptides from
TITLE OF INVENTION: Functional Domains of Bactericidal/Permeability-Increasing
TITLE OF INVENTION: Protein and Uses Thereof
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,344
FILING DATE: 7-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/306,473
FILING DATE: 15-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,752
FILING DATE: 11-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: McDonnell, John J
REGISTRATION NUMBER: 26,949
REFERENCE/DOCKET NUMBER: 93,1133-J
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 910-221-5317
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: "BPI 47"
US-08-473-344-58

Query Match 100.0%; Score 22; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAAK 5

Db   13 KAAAK 17
      |||||

RESULT 10
US-08-473-344-59
; Sequence 59, Application US/08473344
; Patent No. 5763567
; GENERAL INFORMATION:
; APPLICANT: Little, Roger G
; TITLE OF INVENTION: Biologically Active Peptides from
; TITLE OF INVENTION: Functional Domains of Bactericidal/Permeability-Increasing
; TITLE OF INVENTION: Protein and Uses Thereof
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: Illinois
; ZIP: 60606
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0. Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,344
; FILING DATE: 7-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/306,473
; FILING DATE: 15-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/209,762
; FILING DATE: 11-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonnell, John J.
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 93.1133-J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: "BPI.49"
US-08-473-344-59

Query Match   100.0%; Score 22; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 25;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
      |||||
Db   3 KAAAK 7

RESULT 11
PCT-US96-06224-1
; Sequence 1. Application PC/TUS9606224
; GENERAL INFORMATION:
; APPLICANT: POMPLIANO, DAVID L
; APPLICANT: BRAMHILL, DAVID
; APPLICANT: CUNNINGHAM, BARRY R
; APPLICANT: EL'SHERBEINI, MOHANED
; APPLICANT: JONES, BRIAN

; APPLICANT: TRUSCA, DORINA
; TITLE OF INVENTION: A NOVEL BIOCHEMICAL INTERACTION IN
; TITLE OF INVENTION: BACTERIA
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ELLIOTT KORSEN
; STREET: 126 E. LINCOLN AVENUE, P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0. Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/06224
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/437,253
; FILING DATE: 08-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KORSEN, ELLIOTT
; REGISTRATION NUMBER: 32,705
; REFERENCE/DOCKET NUMBER: 19455
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908 594-5493
; TELEFAX: 908 594-4720
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US96-06224-1

Query Match   100.0%; Score 22; DB 5; Length 23;
Best Local Similarity 100.0%; Pred. No. 27;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
      |||||
Db   2 KAAAK 6

RESULT 12
US-09-041-889-41
; Sequence 41, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC pANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,889
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/837,058
FILING DATE: 11-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Campbell Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PM 3006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 55 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-041-889-41

Query Match   100.0%;   Score 22;   DB 3;   Length 55;
Best Local Similarity   100.0%;   Pred. No. 60;
Matches   5;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 KAAAK 5
    | |||| |
Db   7 KAAAK 11

RESULT 13
US-08-346-849-16
; Sequence 16, Application US/08346849
; Patent No. 5670483
GENERAL INFORMATION:
APPLICANT: Zhang, Shuguang
APPLICANT: Lockshin, Curtis
APPLICANT: Rich, Alexander
APPLICANT: Holmes, Todd
TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
TITLE OF INVENTION: THEREFOR
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/346,849
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,326
FILING DATE: 28 DECEMBER 1992
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: MIT-6008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-346-849-16

Query Match   100.0%;   Score 22;   DB 1;   Length 60;
Best Local Similarity   100.0%;   Pred. No. 65;
Matches   5;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 KAAAK 5
    | |||| |
Db   1 KAAAK 5

RESULT 14
US-08-293-284A-16
; Sequence 16, Application US/08293284A
; Patent No. 5955343
GENERAL INFORMATION:
APPLICANT: Holmes, Todd
APPLICANT: Zhang, Shuguang
APPLICANT: Rich, Alexander
APPLICANT: DiPersio, C. Michael
APPLICANT: Lockshin, Curtis
TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
TITLE OF INVENTION: THEREFOR
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/293,284A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,326
FILING DATE: 28-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: MIT-6008A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-293-284A-16

Query Match   100.0%;   Score 22;   DB 2;   Length 60;
Best Local Similarity   100.0%;   Pred. No. 65;
Matches   5;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 KAAAK 5
    | |||| |
Db   1 KAAAK 5

RESULT 15
US-09-041-889-38
; Sequence 38, Application US/09041889
; Patent No. 6033864
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Cohavy, Offer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Ulcerative Colitis and Clinical Subtypes Thereof, Using
; TITLE OF INVENTION: Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Campbell & Flores LLP
;   STREET: 4370 La Jolla Village Drive, Suite 700
;   CITY: San Diego
;   STATE: California
;   COUNTRY: USA
;   ZIP: 92122
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/041,889
;   FILING DATE:
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/837,058
;   FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Campbell, Cathryn A.
;   REGISTRATION NUMBER: 31,815
;   REFERENCE/DOCKET NUMBER: P-PM 3006
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (619) 535-9001
;   TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 38:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 116 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;09-041-889-38.

Query Match   100.0%; Score 22; DB 3; Length 116;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
           |||||
Db 110 KAAAK 114

Search completed: April 24, 2001, 16:36:23
Job time: 303 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:41:56 ; Search time 74.56 seconds
                                          (without alignments)
                                          4.609 Million cell updates/sec

Title: US-09-340-736-4
Perfect score: 22
Sequence: 1 KAAAK 5

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR_67:*
               1: pir1:*
               2: pir2:*
               3: pir3:*
               4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	DB ID	Description
1	22	100.0	18	2	A61220	epsilon receptor m
2	22	100.0	20	2	S06149	photosystem I chai
3	22	100.0	38	2	S23173	photosystem I chai
4	22	100.0	45	2	A05163	antifreeze protein
5	22	100.0	58	2	R71527	probable S21 ribos
6	22	100.0	58	2	P72128	ribosomal protein
7	22	100.0	58	2	B81683	ribosomal protein
8	22	100.0	62	2	T30977	hypothetical prote
9	22	100.0	65	1	R5KT35	ribosomal protein
10	22	100.0	67	2	S76270	ribosomal protein
11	22	100.0	76	2	I45885	elastin - bovine (
12	22	100.0	77	2	S45375	ribosomal protein
13	22	100.0	81	1	JN0462	adenoregulin precu
14	22	100.0	82	2	G75426	hypothetical prote
15	22	100.0	87	2	T27141	hypothetical prote
16	22	100.0	93	2	F70978	hypothetical prote
17	22	100.0	100	2	A33310	nonhistone chromos
18	22	100.0	101	2	S65494	sperm-specific pro
19	22	100.0	105	1	JN0483	ribosomal protein
20	22	100.0	105	2	T08720	ribosomal protein
21	22	100.0	110	2	JC1411	ribosomal protein
22	22	100.0	111	2	S41765	high mobility grou
23	22	100.0	125	1	F1SP4	photosystem I chai
24	22	100.0	125	2	A39396	histone H1-like pr
25	22	100.0	125	2	A71477	probable histone l
26	22	100.0	125	2	B81739	histone H1-like pr
27	22	100.0	132	2	F82800	hypothetical prote
28	22	100.0	136	2	JQ0757	histone H3 staph
29	22	100.0	142	2	C49692	hypothetical prote
30	22	100.0	144	2	T51597	high mobility grou
31	22	100.0	145	2	S53518	histone H2A.2 - wh
32	22	100.0	154	2	JQ1137	rRNA methyltransfe
33	22	100.0	161	2	S18991	high mobility grou
34	22	100.0	162	2	D83911	hypothetical prote
35	22	100.0	164	2	F75184	1su ribosomal prot
36	22	100.0	164	2	F71217	probable ribosomal
37	22	100.0	165	2	A27795	histone H1 - Tetra
38	22	100.0	172	1	D31482	H+-transporting AT
39	22	100.0	173	2	A53729	myosin light chain
40	22	100.0	179	2	G70325	transcription regu
41	22	100.0	182	2	S61926	histone H1 homolog
42	22	100.0	185	2	A34137	histone H1-delta -
43	22	100.0	189	2	S44639	hypothetical prote
44	22	100.0	194	1	HSTRI	histone H1 - trout
45	22	100.0	195	2	E70555	hypothetical prote

ALIGNMENTS

RESULT   1
A61220
epsilon receptor modulating protein (EC 3.4.21.-) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Mar-1996
C;Accession: A61220
R;Matsushita, S.; Katz, D.H.
Cell. Immunol. 137, 252-259, 1991
A;Title: The murine epsilon receptor modulating protein: a novel serine protease whic
A;Reference number: A61220; MUID:91355570
A;Accession: A61220
A;Molecule type: protein
A;Residues: 1-18 <MAY>
C;Comment: This serine proteinase from a T cell hybridoma does not reduce levels of C
C;Keywords: hydrolase; serine proteinase

Query Match   100.0%; Score 22; DB 2; Length 18;
Best Local Similarity   100.0%; Pred. No. 31;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
              |||||
Db   13 KAAAK 17

RESULT   2
S06149
photosystem I chain III - barley (fragment)
C;Species: Hordeum vulgare (barley)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 18-Jun-1993
C;Accession: S06149
R;Anandan, S.; Vainstein, A.; Thornber, J.P.
FEBS Lett. 256, 150-154, 1989
A;Title: Correlation of some published amino acid sequences for photosystem I polypep
A;Reference number: S06149; MUID:90033290
A;Accession: S06149
A;Molecule type: protein
A;Residues: 1-20 <ANA>
A;Note: the authors identified this protein as light-harvesting complex I 17K protein
C;Keywords: chloroplast; photosynthesis; photosystem I; thylakoid

Query Match   100.0%; Score 22; DB 2; Length 20;
Best Local Similarity   100.0%; Pred. No. 34;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
              |||||
Db   12 KAAAK 16

RESULT 3
S23173
photosystem I chain IV precursor - spinach
C;Species: Spinacia oleracea (spinach)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 07-May-1999
C;Accession: S23173
R;Lagoutte, B.; Vallon, O.
A;Title: Purification and membrane topology of PSI-D and PSI-E, two subunits of the phot
Eur. J. Biochem. 205, 1175-1185, 1992
A;Reference number: S23173; MUID:9243534
A;Accession: S23173
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-38 <LAG>

Query Match   100.0%;   Score 22;   DB 2;   Length 38;
Best Local Similarity 100.0%;   Pred. No. 58;
Matches 5;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   1 KAAAK 5
           |||||
Db   20 KAAAK 24

RESULT 4
A05163
antifreeze protein SS-8 - shorthorn sculpin
C;Species: Myoxocephalus scorpius (shorthorn sculpin, daddy sculpin)
C;Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 24-Nov-1999
C;Accession: A05163
R;Hew, C.L.; Joshi, S.; Wang, N.C.; Kao, M.H.; Ananthanarayanan, V.S.
A;Title: Structures of shorthorn sculpin antifreeze polypeptides.
Eur. J. Biochem. 151, 167-172, 1985
A;Reference number: A91150; MUID:85285003
A;Accession: A05163
A;Molecule type: protein
A;Residues: 1-45 <HEW>
C;Superfamily: antifreeze protein
C;Keywords: antifreeze; blocked amino end: plasma
F;9-45/Region: alanine-rich
F;1/Modified site: blocked amino end (Met) #status experimental

Query Match   100.0%;   Score 22;   DB 2;   Length 45;
Best Local Similarity 100.0%;   Pred. No. 66;
Matches 5;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   1 KAAAK 5
           |||||
Db   30 KAAAK 34

RESULT 5
A71527
probable S21 ribosomal protein - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C;Accession: A71527
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A;Reference number: A71570; MUID:99000809
A;Accession: A71527
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-58 <ARN>
A;Cross-references: GB:AE001307; GB:AE001273; NID:g3328757; PIDN:AAC67937.1; PID:g332876
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: rs21

RESULT 4.5
Query Match   100.0%;   Score 22;   DB 2;   Length 58;
Best Local Similarity 100.0%;   Pred. No. 81;
Matches 5;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   1 KAAAK 5
           |||||
Db   50 KAAAK 54

RESULT 6
F72128
ribosomal protein S21 CP0745 [imported] - Chlamydophila pneumoniae (strains CWL029 an
C;Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C;Accession: F72128; G81542
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:99206606
A;Accession: F72128
A;Molecule type: DNA
A;Residues: 1-58 <ARN>
A;Cross-references: GB:AE001588; GB:AE001363; NID:g4376277; PIDN:AAD18184.1; PID:g437
A;Experimental source: strain CWL029
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A;Reference number: G81542; MUID:20150255
A;Accession: G81542
A;Molecule type: DNA
A;Residues: 1-58 <REA>
A;Cross-references: GB:AE002233; GB:AE002216; NID:g7189657; PIDN:AAF38550.1; PID:g718
A;Experimental source: strain AR39, HL cells
C;Genetics:
A;Gene: rs21; CP0745

Query Match   100.0%;   Score 22;   DB 2;   Length 58;
Best Local Similarity 100.0%;   Pred. No. 81;
Matches 5;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   1 KAAAK 5
           |||||
Db   50 KAAAK 54

RESULT 7
B81683
ribosomal protein S21 TC0620 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C;Accession: B81683
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A;Reference number: A81500; MUID:20150255
A;Accession: B81683
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-58 <TET>
A;Cross-references: GB:AE002330; GB:AE002160; NID:g7190649; PIDN:AAF39451.1; PID:g719
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0620

Query Match   100.0%;   Score 22;   DB 2;   Length 58;
Best Local Similarity 100.0%;   Pred. No. 81;
Matches 5;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   1 KAAAK 5

Db   50 KAAAK 54
         |||||

RESULT 8
T30977
hypothetical protein C01B10.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 29-Oct-1999
R;Blanchard, M.; Bradshaw, H.
submitted to the EMBL Data Library, August 1999
A;Description: The sequence of C. elegans cosmid C01B10..
A;Reference number: Z22944
A;Accession: T30977
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
R;Cross-references: EMBL:Z58757; PIDN:AAC47916.1
A;Experimental source: strain Bristol N2
C;Genetics:
A;Map position: IV
A;Introns: 27/3; 61/3
A;Note: C01B10.5

Query Match   100.0%; Score 22; DB 2; Length 62;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAAK 5
      |||||
Db 46 KAAAK 50

RESULT 9
R5KT35
ribosomal protein L35 - Cyanophora paradoxa cyanelle
C;Species: cyanelle Cyanophora paradoxa
C;Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 22-Jun-1999
R;Bryant, D.A.; Stirewalt, V.L.
FEBS Lett. 259, 273-280, 1990
A;Title: The cyanelle genome of Cyanophora paradoxa encodes ribosomal proteins not encod
A;Reference number: S07067; MUID:90092562
A;Accession: S07070
A;Residues: 1-65 <BRY>
R;Cross-references: EMBL:X17063; NID:g11291; PIDN:CAA34907.1; PID:g11292
R;Stirewalt, V.L.; Michalowski, C.B.; Löffelhardt, W.; Bohnert, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A;Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A;Reference number: Z15840
A;Accession: T06849
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
R;Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81192.1; PID:g1016105
A;Experimental source: strain Pringsheim LB555
C;Genetics:
A;Gene: rpl35; rpmI
A;Map position: 27
A;Genome: cyanelle
C;Superfamily: Escherichia coli ribosomal protein L35
C;Keywords: cyanelle; protein biosynthesis; ribosome

Query Match   100.0%; Score 22; DB 1; Length 65;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAAK 5
      |||||

Db   8 KAAAK 12

RESULT 10
S76270
ribosomal protein L35 - Synechocystis sp. (strain PCC 6803)
N;Alternate names: protein ssl1426
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S76270
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S76270
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-67 <KAN>
C;Genetics:
A;Gene: rpl35
C;Superfamily: Escherichia coli ribosomal protein L35
C;Keywords: protein biosynthesis; ribosome

Query Match   100.0%; Score 22; DB 2; Length 67;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAAK 5
      |||||
Db 8 KAAAK 12

RESULT 11
I45885
elastin - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change 13-Aug-1999
C;Accession: I45885
R;Rosenbloom, J.
Lab. Invest. 51, 605-623, 1984
A;Title: Biology of disease: Elastin: Relation of protein and gene structure to disea
A;Reference number: I45885; MUID:85059254
A;Accession: I45885
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Cross-references: GB:M31891; NID:g163008; PIDN:AAA96416.1; PID:g552319
C;Genetics:
A;Introns: 20/1; 58/1
C;Superfamily: elastin

Query Match   100.0%; Score 22; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KAAAK 5
      |||||
Db 12 KAAAK 16

RESULT 12
S45375
ribosomal protein S4 - common tobacco (fragment)
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 06-Jan-1995 #sequence_revision 07-Feb-1997 #text_change 13-Aug-1999
C;Accession: S45375

R;Marty, I.; Brugidou, C.; Chartier, Y.; Meyer, Y.
Plant J. 4, 265-278, 1993
A;Title: Growth-related gene expression in Nicotiana tabacum mesophyll protoplasts.
A;Reference number: S45372; MUID:94035181
A;Accession: S45375
A;Molecule type: mRNA
A;Residues: 1-77 <MAR>
A;Cross-references: EMBL:S14085; NID:g443959; PIDN:CAA78463.1; PID:g443960
C;Superfamily: rat ribosomal protein S9
C;Keywords: protein biosynthesis; ribosome

```
Query Match    100.0%;  Score 22;  DB 2;  Length 77;
Best Local Similarity  100.0%;  Pred. No. 1e+02;
Matches    5;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 KAAAK 5
           |||||
Db    60 KAAAK 64
```

RESULT 13
G
162
Adoregulin precursor - two-colored leaf frog
N;Alternate names: dermaseptin b II
C;Species: Phyllomedusa bicolor (two-colored leaf frog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: JN0462; A54897; A44171; S34718
R;Amiche, M.; Ducancel, F.; Lajeunesse, E.; Boulain, J.C.; Menez, A.; Nicolas, P.
Biochem. Biophys. Res. Commun. 191, 983-990, 1993
A;Title: Molecular cloning of a cDNA encoding the precursor of adenoregulin from frog sk
A;Reference number: JN0462; MUID:93221546
A;Accession: JN0462
A;Molecule type: mRNA
A;Residues: 1-81 <AMI>
A;Experimental source: skin
A;Cross-references: EMBL:X70278; NID:g395931; PIDN:CAA49763.1; PID:g395932
A;Note: the authors translated the codon AAA for residue 26 as Leu
R;Amiche, M.; Ducancel, F.; Mor, A.; Boulain, J.C.; Menez, A.; Nicolas, P.
J. Biol. Chem. 269, 17847-17852, 1994
A;Title: Precursors of vertebrate peptide antibiotics dermaseptin b and adenoregulin hav
A;Reference number: A54897; MUID:94299491
A;Accession: A54897
A;Molecule type: mRNA
A;Residues: 1-81 <AM2>
A;Cross-references: GB:X70278; NID:g395931; PIDN:CAA49763.1; PID:g395932
R;Daly, J.W.; Caceres, J.; Moni, R.W.; Guspvsky, F.; Moos Jr., M.; Seamon, K.B.; Milton,
Proc. Natl. Acad. Sci. U.S.A. 89, 10960-10963, 1992
A;Title: Frog secretions and hunting magic in the upper Amazon: identification of a pept
ference number: A44171; MUID:93065363
A;Molecule type: protein
A;Residues: 46-78 <DAL>
C;Superfamily: dermaseptin precursor; dermorphin precursor amino-terminal homology
C;Keywords: antibiotic; antifungal; skin
F;1-45/Domain: dermorphin precursor amino-terminal homology <DER>
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-45/Domain: propeptide #status predicted <PRO>
F;46-78/Product: adenoregulin #status experimental <MAT>

```
Query Match    100.0%;  Score 22;  DB 1;  Length 81;
Best Local Similarity  100.0%;  Pred. No. 1e+02;
Matches    5;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 KAAAK 5
           |||||
Db    60 KAAAK 64
```

RESULT 14
G75426

hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: G75426
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
. M.; Shen, M.; Yamachevan, J.J.; Lam, P.; McDonald, L.; Otterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1
A;Reference number: A75250; MUID:20036896
A;Accession: G75426
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-82 <WHI>
A;Cross-references: GB:AE001867; GB:AE000513; NID:g6458915; PIDN:AAF10756.1; PID:g645
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1176
A;Map position: 1

```
Query Match    100.0%;  Score 22;  DB 2;  Length 82;
Best Local Similarity  100.0%;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
Matches    5;

Qy     1 KAAAK 5
           |||||
Db    62 KAAAK 66
```

RESULT 15
T27141
Hypothetical protein Y53F4A.b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27141
R;Smye, R.
submitted to the EMBL Data Library, September 1999
A;Reference number: Z20318
A;Accession: T27141
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-87 <WIL>
A;Cross-references: EMBL:AL110494; NID:e1542292; PIDN:CAB54462.1; CESP:Y53F4A.b
A;Experimental source: clone Y53F4A
C;Genetics:
A;Gene: CESP:Y53F4A.b
A;Introns: 76/3

```
Query Match    100.0%;  Score 22;  DB 2;  Length 87;
Best Local Similarity  100.0%;  Pred. No. 1.1e+02;
Matches    5;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 KAAAK 5
           |||||
Db    26 KAAAK 30
```

Search completed: April 24, 2001, 16:41:58
Job time: 463 sec

us-09-340-736-4.rpr

```
               GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     April 24, 2001, 16:42:47 ; Search time 44.88 Seconds
                                        (without alignments)
                                        3.816 Million cell updates/sec

Title:           US-09-340-736-4
Perfect score:   22
Sequence:        1 KAAAK 5

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Matched:         93435 seqs. 34255486 residues

Total number of hits satisfying chosen parameters:     93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

   Pred. No. is the number of results predicted by chance to have a
   score greater than or equal to the score of the result being printed,
   and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	22	100.0	27	1	DMS4_PHYSA	P80280 phyllomedus
2	22	100.0	29	1	DMS5_PHYSA	P80281 phyllomedus
3	22	100.0	45	1	ANP8_MYOSC	P04368 myoxocephal
4	22	100.0	58	1	RS21_CHLPN	Q9Z9F0 chlamydia p
5	22	100.0	58	1	RS21_CHLTR	O84346 chlamydia t
6	22	100.0	65	1	PRK35_CYAPA	P14810 cyanophora
7	22	100.0	67	1	RL35_SYNY3	P48959 synechocyst
8	22	100.0	74	1	RIC8_TRYCR	P40270 trypanosoma
9	22	100.0	77	1	RS9_TOBAC	P19214 nicotiana t
10	22	100.0	80	1	RL61_TRYCR	P40273 trypanosoma
11	22	100.0	81	1	DMS2_PHYBI	P31107 phyllomedus
12	22	100.0	90	1	H2162_TRYCR	P40274 trypanosoma
13	22	100.0	99	1	HG14_HUMAN	P65114 homo sapien
14	22	100.0	104	1	RL36_HUMAN	Q9Y3U8 homo sapien
15	22	100.0	104	1	RL36_RAT	P39032 rattus norv
16	22	100.0	110	1	RS25_DICDI	O03409 dictyosteli
17	22	100.0	111	1	HMG2_DROME	O06943 drosophila
18	22	100.0	125	1	HCT1_CHLTR	O02281 chlamydia t
19	22	100.0	125	1	PSAE_SPIOL	P12354 spinacia ol
20	22	100.0	135	1	H3_ACRFO	P22843 acropora fo
21	22	100.0	154	1	KAMB_STRTN	P25920 streptomyce
22	22	100.0	161	1	HMGL_WHEAT	P40621 triticum ae
23	22	100.0	164	1	RL11_PYRAB	Q9V2S3 pyrococcus
24	22	100.0	164	1	RL11_PYRHO	O57779 pyrococcus
25	22	100.0	165	1	RL21_ARATH	O43291 arabidopsis
26	22	100.0	166	1	HL_TETPY	P12305 tetrahymena
27	22	100.0	172	1	RL14_DROME	P55841 drosophila
28	22	100.0	185	1	ATPB_BACME	P20601 bacillus me
29	22	100.0	189	1	HID_STRPU	P15870 strongyloce
30	22	100.0	194	1	YFP2_CAEEL	P41800 caenorhabdi
31	22	100.0	197	1	HL_SALTR	P02254 salmo trutt
32	22	100.0	199	1	RL22_MYCTU	P95054 mycobacteri
33	22	100.0		1	Y475_MYCTU	Q11142 mycobacteri
34	22	100.0	205	1	H1E_STRPU	P19375 strongyloce
35	22	100.0	206	1	H1_ONCMY	P06350 oncorhynchu
36	22	100.0	209	1	H1A_XENLA	P06892 xenopus lae
37	22	100.0	211	1	H1B_STRPU	P15869 strongyloce
38	22	100.0	213	1	DBH_MYCTU	P02251 oryctolagus
39	22	100.0	214	1	H1CL_XENLA	P95109 mycobacteri
40	22	100.0	216	1	H1G_STRPU	P15866 xenopus lae
41	22	100.0	217	1	H1_CHICK	P07796 strongyloce
42	22	100.0	217	1	H1O1_CHICK	P09987 gallus gall
43	22	100.0	218	1	H1R_CHICK	P08284 gallus gall
44	22	100.0	218	1	H1B_HUMAN	P08288 gallus gall
45	22	100.0	218	1		P10412 homo sapien

```
                              ALIGNMENTS

RESULT  1
DMS4_PHYSA
ID   DMS4_PHYSA    STANDARD;      PRT;   27 AA.
AC   P80280;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DERMASEPTIN 4 (DS IV)
OS   Phyllomedusa sauvagei (Sauvage's leaf frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;
OC   Phyllomedusa.
OX   NCBI_TaxID=8395;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Skin;
RX   MEDLINE=94139686; PubMed=8306981;
RA   Mor A., Nicolas P.;
RT   "Isolation and structure of novel defensive peptides from frog skin.";
RL   Eur. J. Biochem. 219:145-154(1994).
CC   -!- FUNCTION: POSSESSES A POTENT ANTIMICROBIAL ACTIVITY AGAINST
CC       BACTERIA FUNGI AND PROTOZOA. PROBABLY ACTS BY DISTURBING MEMBRANE
CC       FUNCTIONS WITH ITS AMPHIPATIC STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: SKIN.
KW   Antibiotic; Fungicide; Multigene family; Amphibian skin.
SQ   SEQUENCE   27 AA;  2779 MW;  43C94D2DC19721A8 CRC64;

Query Match             100.0%;  Score 22;  DB 1;  Length 27;
Best Local Similarity   100.0%;            Pred. No. 17;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 KAAAK 5
                |||||
Db         12 KAAAK 16


RESULT  2
DMS5_PHYSA
ID   DMS5_PHYSA    STANDARD;      PRT;   29 AA.
AC   P80281;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DERMASEPTIN 5 (DS V).
OS   Phyllomedusa sauvagei (Sauvage's leaf frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;
OC   Phyllomedusa.
OX   NCBI_TaxID=8395;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Skin;
RX   MEDLINE=94139686; PubMed=8306981;
RA   Mor A., Nicolas P.;
```

RT   "Isolation and structure of novel defensive peptides from frog skin.";
RL   Eur. J. Biochem. 219.145-154(1994).
CC   -!- FUNCTION: POSSESSES A POTENT ANTIMICROBIAL ACTIVITY AGAINST
CC   BACTERIA FUNGI AND PROTOZOA. PROBABLY ACTS BY DISTURBING MEMBRANE
CC   FUNCTIONS WITH ITS AMPHIPATIC STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: SKIN.
KW   Antibiotic; Fungicide; Multigene family; Amphibian skin.
SQ   SEQUENCE   29 AA; 2840 MW; 540A4971FC5BB506 CRC64;

Query Match   100.0%; Score 22; DB 1; Length 29;
Best Local Similarity   100.0%; Pred. No. 18;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
              |||||
Db   15 KAAAK 19

RESULT 3
ANP8_MYOSC
ID   ANP8_MYOSC   STANDARD;   PRT;   45 AA.
AC   P04368;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   01-AUG-1990 (Rel. 15, Last annotation update)
DE   ANTIFREEZE PEPTIDE SS-8.
OS   Myoxocephalus scorpius (Shorthorn sculpin) (Daddy sculpin).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;
OC   Cottoidei; Cottidae; Myoxocephalus.
OX   NCBI_TaxID=8097;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=85285603; PubMed=4029130;
RA   Hew C.-L., Joshi S., Wang N.-C., Kao M.H., Ananthanarayanan V.S.;
RT   "Structures of shorthorn sculpin antifreeze polypeptides.";
RL   Eur. J. Biochem. 151:167-172(1985).
CC   -!- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC   -!- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE
CC   ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
DR   PIR; A05163; A05163.
DR   HSSP; P04002; 1WFA.
KW   Antifreeze protein; Repeat.
FT   MOD_RES   1   1   BLOCKED.
FT   REPEAT   9   21   Repeat.
FT   REPEAT   22   33
FT   REPEAT   34   45
SQ   SEQUENCE   45 AA; 4006 MW; 260C0BCC663B6878 CRC64;

Query Match   100.0%; Score 22; DB 1; Length 45;
Best Local Similarity   100.0%; Pred. No. 27;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
              |||||
Db   30 KAAAK 34

RESULT 4
RS21_CHLPN
ID   RS21_CHLPN   STANDARD;   PRT;   58 AA.
AC   Q929F0; Q9Z0K5;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   30S RIBOSOMAL PROTEIN S21.
GN   RPSU OR RS21 OR CPN031 OR CP0745.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.

OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TW183;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
CC   -!- SIMILARITY: BELONGS TO THE S21P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AE001588; AAD18184.1; -.
DR   EMBL; AE002233; AAF38550.1; -.
DR   EMBL; AE002545; BAA98243.1; -.
DR   TIGR; CP0745; -.
DR   InterPro; IPR001911; -.
DR   Pfam; PF01165; Ribosomal_S21; 1.
DR   PRINTS; PR00976; RIBOSOMALS21.
DR   PROSITE; PS01181; RIBOSOMAL_S21; FALSE_NEG.
KW   Ribosomal protein.
SQ   SEQUENCE   58 AA; 6636 MW; 74118A525600177 CRC64;

Query Match   100.0%; Score 22; DB 1; Length 58;
Best Local Similarity   100.0%; Pred. No. 34;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
              |||||
Db   50 KAAAK 54

RESULT 5
RS21_CHLTR
ID   RS21_CHLTR   STANDARD;   PRT;   58 AA.
AC   O84346;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   30S RIBOSOMAL PROTEIN S21.
GN   RPSU OR RS21 OR CT342 OR TC0620.
OS   Chlamydia trachomatis, and
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX   NCBI_TaxID=813, 83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=C.trachomatis; STRAIN=D/UW-3/CX;
RX   MEDLINE=99000809; PubMed=9784136;
RA   Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA   Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA   Davis R.W.;
RT   "Genome sequence of an obligate intracellular pathogen of humans:
RT   Chlamydia trachomatis.";
RL   Science 282:754-759(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=C.muridarum; STRAIN=MOPN / NIGG;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K.,
RA   Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
RA   Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G.,
RA   Salzberg S.L., Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
CC   -!- SIMILARITY: BELONGS TO THE S21P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AE001307; AAC67937.1; -.
DR   EMBL; AE002330; AAF39451.1; -.
DR   TIGR; TC0620; -.
DR   InterPro; IPR001911; -.
DR   Pfam; PF01165; Ribosomal_S21; 1.
DR   PROSITE; PS01181; RIBOSOMAL_S21; FALSE_NEG.
KW   Ribosomal protein.
SQ   SEQUENCE   58 AA;   6666 MW;   99AD89D046251852 CRC64;

Query Match   100.0%; Score 22; DB 1; Length 58;
Best Local Similarity   100.0%; Pred. No. 34;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
            |||||
Db   50 KAAAK 54

RESULT 6
RK35_CYAPA
ID   RK35_CYAPA   STANDARD;   PRT;   65 AA.
AC   P14810;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   CYANELLE 50S RIBOSOMAL PROTEIN L35.
GN   RPL35.
OS   Cyanophora paradoxa.
OC   Cyanelle.
OC   Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX   NCBI_TaxID=2762;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LB555 / PRINGSHEIM;
RX   MEDLINE=90092562; PubMed=2403527;
RA   Bryant D.A., Stirewalt V.L.;
RT   "The cyanelle genome of Cyanophora paradoxa encodes ribosomal
RT   proteins not encoded by the chloroplasts genomes of higher plants.";
RL   FEBS Lett. 259:273-280(1990).

RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LB555 / PRINGSHEIM;
RA   Stirewalt V.L., Michalowski C.B., Luffelhardt W., Bohnert H.J.,
RA   Bryant D.A.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE L35P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X17063; CAA34907.1; -.
DR   EMBL; U30821; AAA81192.1; -.
DR   PIR; S07070; R5KT35.
DR   Mendel; 7912; CYApa:rpl35.1.
DR   InterPro; IPR001706; -.
DR   Pfam; PF01632; Ribosomal_L35p; 1.
DR   PRINTS; PR00064; RIBOSOMALL35.
DR   PROSITE; PS00936; RIBOSOMAL_L35; 1.
KW   Ribosomal protein; Cyanelle.
SQ   SEQUENCE   55 AA;   7607 MW;   D689E2C85CA95B4F CRC64;

Query Match   100.0%; Score 22; DB 1; Length 65;
Best Local Similarity   100.0%; Pred. No. 37;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
            |||||
Db   8 KAAAK 12

RESULT 7
RL35_SYNY3
ID   RL35_SYNY3   STANDARD;   PRT;   57 AA.
AC   P48959;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L35.
GN   RPMI OR RPL35 OR SSL1426.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96127529; PubMed=8590279;
RA   Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA   Sugiura M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC5803. I. Sequence features in the 1 Mb
RT   region from map positions 64% to 92% of the genome.";
RL   DNA Res. 2:153-166(1995).
CC   -!- SIMILARITY: BELONGS TO THE L35P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; D64000; BAA10122.1; -.
DR   InterPro; IPR001706; -.
DR   Pfam; PF01632; Ribosomal_L35p; 1.
DR   PRINTS; PR00064; RIBOSOMALL35.
DR   PROSITE; PS00936; RIBOSOMAL_L35; 1.

KW   Ribosomal protein.
SQ   SEQUENCE   67 AA;   7891 MW; E43BB0E79BA5F294 CRC64;

Query Match   100.0%; Score 22; DB 1; Length 67;
Best Local Similarity 100.0%; Pred. No. 38;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
              |||||
Db   8 KAAAK 12

RESULT 8
H1C8_TRYCR
ID   H1C8_TRYCR   STANDARD;   PRT;   74 AA.
AC   P40270; P40272;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HISTONE H1.C8/H1.M1.
OS   Trypanosoma cruzi.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5693;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TULAHUEN 2;
RX   MEDLINE=95059220; PubMed=7969272;
RA   Aslund L., Carlsson L., Henriksson J., Rydaaker M., Toro G.C.,
RA   Galanti N., Pettersson U.;
RT   "A gene family encoding heterogeneous histone H1 proteins in
RT   Trypanosoma cruzi.";
RL   Mol. Biochem. Parasitol. 65:317-330(1994).
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; L27116; AAA65480.1; -.
DR   EMBL; L27118; AAA66482.1; -.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ   SEQUENCE   74 AA;   7668 MW; 036AF6C72BB0249B CRC64;

Query Match   100.0%; Score 22; DB 1; Length 74;
Best Local Similarity 100.0%; Pred. No. 42;
ches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
              |||||
Db   15 KAAAK 19

RESULT 9
RS9_TOBAC
ID   RS9_TOBAC   STANDARD;   PRT;   77 AA.
AC   P49214;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   40S RIBOSOMAL PROTEIN S9 (S4) (FRAGMENT).
GN   RPS9
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC   Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. 19;
RX   MEDLINE=94035181; PubMed=8220482;
RA   Marty I., Brugidou C., Chartier Y., Meyer Y.;
RT   "Growth-related gene expression in Nicotiana tabacum mesophyll
RT   protoplasts.";
RL   Plant J. 4:265-278(1993).
CC   -!- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; Z14085; CAA78463.1; -.
DR   InterPro; IPR001912; -.
DR   InterPro; IPR000942; -.
DR   Pfam; PF01479; S4; 1.
DR   PROSITE; PS00632; RIBOSOMAL_S4; PARTIAL.
KW   Ribosomal protein.
FT   NON_TER   1   1
SQ   SEQUENCE   77 AA;   8557 MW; 07EE8031A7D25B2 CRC64;

Query Match   100.0%; Score 22; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 43;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
              |||||
Db   60 KAAAK 64

RESULT 10
H161_TRYCR
ID   H161_TRYCR   STANDARD;   PRT;   80 AA.
AC   P40273;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HISTONE H1.M6.1.
OS   Trypanosoma cruzi.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5693;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TULAHUEN 2;
RX   MEDLINE=95059220; PubMed=7969272;
RA   Aslund L., Carlsson L., Henriksson J., Rydaaker M., Toro G.C.,
RA   Galanti N., Pettersson U.;
RT   "A gene family encoding heterogeneous histone H1 proteins in
RT   Trypanosoma cruzi.";
RL   Mol. Biochem. Parasitol. 55:317-330(1994).
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; L27119; AAA66483.1; -.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ   SEQUENCE   80 AA;   8271 MW; 8E83DA508F0A5D31 CRC64;

Query Match   100.0%; Score 22; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 45;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy      1 KAAAK 5
          |||||
Db     15 KAAAK 19


RESULT 11
DM52_PHYBI
ID   DM52_PHYBI          STANDARD;      PRT;     81 AA.
AC   P31107; P80283;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   ADENOREGULIN PRECURSOR (DERMASEPTIN BII).
OS   Phyllomedusa bicolor (Two-colored leaf frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;
OC   Phyllomedusa.
OX   NCBI_TaxID=8393;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=93221546; PubMed=84660537;
RA   Amiche M., Ducancel F., Lajennesse E., Boulain J.-C., Menez A.,
RA   Nicolas P.;
RT   "Molecular cloning of a cDNA encoding the precursor of adenoregulin
RT   from frog skin. Relationships with the vertebrate defensive peptides,
RT   dermaseptins.";
RL   Biochem. Biophys. Res. Commun. 191:983-990(1993).
RN   [2]
RP   SEQUENCE OF 46-78.
RC   TISSUE=Skin;
RX   MEDLINE=93066363; PubMed=1438301;
RA   Daly J.W., Caceres J., Moni R.W., Gusovsky F., Moos M. Jr.,
RA   Seamon K.B., Milton K., Myers C.W.;
RT   "Frog secretions and hunting magic in the upper Amazon:
RT   identification of a peptide that interacts with an adenosine
RT   receptor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:10960-10963(1992).
RN   [3]
RP   SYNTHESIS, AND CHARACTERIZATION.
RC   TISSUE=Skin;
RX   MEDLINE=94139686; PubMed=8306981;
RA   Mor A., Nicolas P.;
RT   "Isolation and structure of novel defensive peptides from frog skin.";
RL   Eur. J. Biochem. 219:145-154(1994).
CC   -!- FUNCTION: ENHANCES BINDING OF AGONISTS TO A1 ADENOSINE RECEPTORS.
CC   -!- FUNCTION: POSSESSES A POTENT ANTIMICROBIAL ACTIVITY AGAINST
CC       BACTERIA, FUNGI AND PROTOZOA. PROBABLY ACTS BY DISTURBING MEMBRANE
CC       FUNCTIONS WITH ITS AMPHIPATHIC STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: SKIN.
CC   -!- DISEASE: AFFECTS HUMAN BEHAVIOR ELICITING PROFOUND MALAISE,
CC       FOLLOWED BY LISTLESSNESS AND THEN EUPHORIA.
CC   -!- SIMILARITY: TO P. SAUVAGEI DERMASEPTIN.
CC       ---------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       ---------------------------------------------------------------------------
DR   EMBL; X70278; CAA49763.1; -.
DR   PIR; A44171; A44171.
DR   PIR; JN0462; JN0462.
KW   Antibiotic; Fungicide; Multigene family; Amphibian skin; Signal;
KW   Cleavage on pair of basic residues.
FT   SIGNAL        1     22       POTENTIAL.
FT   PROPEP       23     45       POTENTIAL.
FT   CHAIN        46     78       ADENOREGULIN.
FT   PROPEP       79     81       POTENTIAL.
SQ   SEQUENCE   81 AA;  8844 MW;  C26AD84B941B272D CRC64;
```

```
Query Match              100.0%;  Score 22;  DB 1;  Length 81;
Best Local Similarity    100.0%;  Pred. No. 45;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KAAAK 5
          |||||
Db     60 KAAAK 64


RESULT 12
H162_TRYCR
ID   H162_TRYCR          STANDARD;      PRT;     90 AA.
AC   P40274;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HISTONE H1.M6.2.
OS   Trypanosoma cruzi.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5693;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TULAHUEN 2;
RX   MEDLINE=95059220; PubMed=7969272;
RA   Aaslund L., Carlsson L., Henriksson J., Rydaaker M., Toro G.C.,
RA   Galanti N., Pettersson U.;
RT   "A gene family encoding heterogeneous histone H1 proteins in
RT   Trypanosoma cruzi.";
RL   Mol. Biochem. Parasitol. 65:317-330(1994).
CC       ---------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       ---------------------------------------------------------------------------
DR   EMBL; L27119; AAA66484.1; -.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ   SEQUENCE   90 AA;  9236 MW;  B99950A99D598D6C CRC64;
```

```
Query Match              100.0%;  Score 22;  DB 1;  Length 90;
Best Local Similarity    100.0%;  Pred. No. 50;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KAAAK 5
          |||||
Db     15 KAAAK 19


RESULT 13
HG14_HUMAN
ID   HG14_HUMAN          STANDARD;      PRT;     99 AA.
AC   P05114;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   NONHISTONE CHROMOSOMAL PROTEIN HMG-14.
GN   HMG14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89123472; PubMed=2561381;
RA   Landsman D., McBride O.W., Soares N., Crippa M.P., Srikantha T.,
RA   Bustin M.;
RT   "Chromosomal protein HMG-14. Identification, characterization, and
```

RT   chromosome localization of a functional gene from the large human
RT   multigene family.";
RL   J. Biol. Chem. 264:3421-3427(1989).
RN   [2]
RX   MEDLINE=87057273; PubMed=3782107;
RA   Landsman D., Srikantha T., Westermann R., Bustin M.;
RT   "Chromosomal protein HMG-14. Complete human cDNA sequence and
RT   evidence for a multigene family";
RL   J. Biol. Chem. 261:16082-16086(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Rump A., Blechschmidt K., Dagand E., Hildmann T., Nordsiek G.,
RA   Dresche B., Weber J., Schatteroy R., Yaspo M.-L., Rosenthal A.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schilhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
CC   -!- FUNCTION: BIND TO THE INNER SIDE OF THE NUCLEOSOMAL DNA THUS
CC   ALTERING THE INTERACTION BETWEEN THE DNA AND THE HISTONE OCTAMER.
CC   MAY BE INVOLVED IN THE PROCESS WHICH MAINTAINS TRANSCRIBABLE GENES
CC   IN A UNIQUE CHROMATIN CONFORMATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE HMG-14/HMG-17 PROTEIN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M21339; AAA52677.1; -.
DR   EMBL; J02621; AAA52676.1; -.
DR   EMBL; AF064861; AAC18356.1; -.
DR   EMBL; AL163279; CAB90453.1; -.
DR   PIR; A26326; A26326.
DR   PIR; A33310; A33310.
DR   MIM; 163920; -.
DR   InterPro; IPR000079; -.
DR   Pfam; PF01101; HMG14_17; 1.
DR   PRINTS; PR00925; NONHISHMG17.
DR   PROSITE; PS00355; HMG14_17; 1.
KW   Nuclear protein; DNA-binding.
FT   INIT_MET   0   0
FT   CONFLICT   12   12   MISSING (IN REF. 2).
SQ   SEQUENCE   99 AA; 10528 MW; 4CD91C57B4C162FA CRC64;

Query Match   100.0%; Score 22; DB 1; Length 99;
Best Local Similarity   100.0%; Pred. No. 54;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
          |||||
Db   37 KAAAK 41

RESULT 14
RL36_HUMAN
ID   RL36_HUMAN   STANDARD;   PRT;   104 AA.
AC   Q9Y3U8;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   60S RIBOSOMAL PROTEIN L36.
GN   RPL36.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE L36E FAMILY OF RIBOSOMAL PROTEINS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AL050273; CAB43374.1; -.
DR   InterPro; IPR000509; -.
DR   Pfam; PF01158; Ribosomal_L36e; 1.
DR   PROSITE; PS01190; RIBOSOMAL_L36E; 1.
KW   Ribosomal protein.
FT   INIT_MET   0   0   BY SIMILARITY.
SQ   SEQUENCE   104 AA; 12122 MW; A47D76C55EFB9B1A CRC64;

Query Match   100.0%; Score 22; DB 1; Length 104;
Best Local Similarity   100.0%; Pred. No. 57;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
          |||||
Db   98 KAAAK 102

RESULT 15
RL36_RAT
ID   RL36_RAT   STANDARD;   PRT;   104 AA.
AC   P39032;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   60S RIBOSOMAL PROTEIN L36.
GN   RPL36.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-39.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX   MEDLINE=93249466; PubMed=8484789;
RA   Chan Y.-L., Paz V., Olvera J., Wool I.G.;
RT   "The primary structure of rat ribosomal protein L36.";
RL   Biochem. Biophys. Res. Commun. 192:849-853(1993).
CC   -!- SIMILARITY: BELONGS TO THE L36E FAMILY OF RIBOSOMAL PROTEINS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial

CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X68234; CAA48345.1; -.
DR   PIR; JN0483; JN0483.
DR   InterPro; IPR000509; -.
DR   Pfam; PF01158; Ribosomal_L36e; 1.
DR   PROSITE; PS01190; RIBOSOMAL_L36E; 1.
KW   Ribosomal protein.
FT   INIT_MET   0   0
SQ   SEQUENCE   104 AA;   12136 MW; A27B76C358FB9B1A CRC64;

Query Match   100.0%; Score 22; DB 1; Length 104;
Best Local Similarity   100.0%; Pred. No. 57;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
              |||||
Db   98 KAAAK 102

Search completed: April 24, 2001, 16:42:48
Job time: 454 sec

OM protein - protein search, using sw model

Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5

Run on: April 24, 2001, 16:40:30 ; Search time 125.5 Seconds
                                    (without alignments)
                                    4.670 Million cell updates/sec

Title: US-09-340-736-4
Perfect score: 22
Sequence: 1 KAAAK 5

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

...ched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database ;   SPTREMBL_15:*
        1:   sp_archea:*
        2:   sp_bacteria:*
        3:   sp_fungi:*
        4:   sp_human:*
        5:   sp_invertebrate:*
        6:   sp_mammal:*
        7:   sp_mhc:*
        8:   sp_organelle:*
        9:   sp_phage:*
       10:   sp_plant:*
       11:   sp_rodent:*
       12:   sp_unclassified:*
       13:   sp_vertebrate:*
       14:   sp_virus:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	22	100.0	15	10	Q9SQI6	Q9sqi6 oryza sativ
2	22	100.0	24	10	Q9S825	Q9s825 spinacia ol
3	22	100.0	45	2	Q01215	rhizomucor
4	22	100.0	58	2	Q9JQK5	chlamydia p
5	22	100.0	60	5	Q9U3V3	caenorhabdi
6	22	100.0	61	5	Q9NFJ7	Q9nfj7 trypanosoma
7	22	100.0	61	5	Q9NFJ5	Q9nfj5 trypanosoma
8	22	100.0	62	5	Q17536	caenorhabdi
9	22	100.0	66	5	Q9NFJ3	Q9nfj3 trypanosoma
10	22	100.0	71	5	Q9NFJ8	Q9nfj8 trypanosoma
11	22	100.0	71	5	Q9N6K5	Q9n6k5 trypanosoma
12	22	100.0	71	5	Q9N6K0	Q9n6k0 trypanosoma
13	22	100.0	72	2	P96100	thiobacillu
14	22	100.0	75	4	Q9NFK0	Q9nfk0 trypanosoma
15	22	100.0	76	4	Q9Y5V0	Q9y5v0 homo sapien
16	22	100.0	76	5	Q9N6J6	Q9nfj6 trypanosoma
17	22	100.0	76	5	Q9NJ74	Q9nfj4 trypanosoma
18	22	100.0	76	5	Q9NFF2	Q9nff2 trypanosoma
19	22	100.0	76	5	Q9NFF1	Q9nffi trypanosoma
20	22	100.0	76	5	Q9N6N7	trypanosoma
21	22	100.0	76	5	Q9N6J8	trypanosoma
22	22	100.0	76	5	Q9N6J0	trypanosoma
23	22	100.0	76	6	Q28100	bos taurus
24	22	100.0	80	5	Q9N6S6	trypanosoma
25	22	100.0	80	13	Q93453	pachymedusa
26	22	100.0	81	5	Q9NFG6	trypanosoma
27	22	100.0	82	5	Q9NCL9	deinococcus
28	22	100.0	82	5	Q9N6S4	trypanosoma
29	22	100.0	85	5	Q26577	trypanosoma
30	22	100.0	85	5	Q9N6G8	trypanosoma
31	22	100.0	87	5	Q9NAC9	caenorhabdi
32	22	100.0	91	5	Q9NFJ9	trypanosoma
33	22	100.0	93	2	P96874	mycobacteri
34	22	100.0	101	5	Q9TXE3	mytilus tro
35	22	100.0	105	4	Q9Q0F6	homo sapien
36	22	100.0	112	5	Q9XYY5	leishmania
37	22	100.0	125	2	Q9PII1	chlamydia m
38	22	100.0	132	2	Q9PG09	xylella fas
39	22	100.0	136	2	P72462	streptoallo
40	22	100.0	137	10	Q9SHE5	arabidopsis
41	22	100.0	142	2	Q02797	streptomyce
42	22	100.0	144	10	O49596	arabidopsis
43	22	100.0	145	10	Q43213	triticum ae
44	22	100.0	149	8	Q43346	unidentifie
45	22	100.0	162	2	Q9KB42	bacillus ha

ALIGNMENTS

RESULT 1
Q9SQI6
ID   Q9SQI6;   PRELIMINARY;   PRT;   15 AA.
AC   Q9SQI6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   TRANSLATION ELONGATION FACTOR (FRAGMENT).
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Li Z.-Y., Chen S.-Y.;
RT   "Inducible expression of translation elongation factor 1A gene in rice
RT   seedlings in response to environmental stresses.";
RL   Acta Bot. Sin. 41:800-806(1999). -.
DR   EMBL; AF067195; AAC79991.1; -.
KW   Elongation factor;
FT   NON_TER   1   1
SQ   SEQUENCE   15 AA;   1514 MW; 95E5C46F069C9775 CRC64;

Query Match   100.0%; Score 22; DB 10; Length 15;
Best Local Similarity   100.0%; Pred. No. 28;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 KAAAK 5
            |||||
Db   9 KAAAK 13

RESULT 2
Q9S825
ID   Q9S825;   PRELIMINARY;   PRT;   24 AA.
AC   Q9S825;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PHOTOSYSTEM I REACTION CENTER SUBUNIT IV, PSI-E.
OS   Spinacia oleracea (Spinach).

```
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;
OC   Caryophyllales; Chenopodiaceae; Spinacia.
OX   NCBI_TaxID=3562;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=92249324; PubMed=1374333;
RA   Lagoutte B., Vallon O.;
RT   "Purification and membrane topology of PSI-D and PSI-E, two subunits
RT   of the photosystem I reaction center.";
RL   Eur. J. Biochem. 205.1175-1185(1992).
SQ   SEQUENCE   24 AA;  2150 MW;  5CD2898B5B18955B CRC64;

Query Match             100.0%;  Score 22;  DB 10;  Length 24;
Best Local Similarity   100.0%;  Pred. No. 43;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KAAAK 5
           |||||
Db      20 KAAAK 24

RESULT 3
Q01215
ID   Q01215        PRELIMINARY;       PRT;   45 AA.
AC   Q01215; Q01216;
DT   01-NOV-1996 (TREMBLrel. 01, Created)
DT   01-JAN-1999 (TREMBLrel. 09, Last sequence update)
DT   01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE   ELONGATION FACTOR 1-ALPHA (EF-2) (FRAGMENT)
OS   Rhizomucor racemosus (Mucor circinelloides f. lusitanicus).
OC   Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC   Mucor.
OX   NCBI_TaxID=4841;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87257895; PubMed=3037339;
RA   Linz J.E., Sypherd P.S.;
RT   "Expression of three genes for elongation factor 1 alpha during
RT   morphogenesis of Mucor racemosus.";
RL   Mol. Cell. Biol. 7:1925-1932(1987).
DR   EMBL; M16353; AAA33426.1; -.
DR   EMBL; M16354; AAA33428.1; -.
DR   MENDEL; 20745; Rhira;3109;20745.
KW   Elongation factor.
FT   NON_TER       1      1
SQ   SEQUENCE   45 AA;  4801 MW;  C3E98BD03229AD80 CRC64;

Query Match             100.0%;  Score 22;  DB 3;  Length 45;
Best Local Similarity   100.0%;  Pred. No. 78;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KAAAK 5
           |||||
Db      37 KAAAK 41

RESULT 4
Q9JQK5
ID   Q9JQK5        PRELIMINARY;       PRT;   58 AA.
AC   Q9JQK5;
DT   01-OCT-2000 (TREMBLrel. 15, Created)
DT   01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE   S21 RIBOSOMAL PROTEIN.
GN   RS21.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28.2311-2314(2000).
DR   EMBL; AP002545; BAA98243.1; -.
KW   Ribosomal protein.
SQ   SEQUENCE   58 AA;  6636 MW;  7411BAE52560D177 CRC64;

Query Match             100.0%;  Score 22;  DB 2;  Length 58;
Best Local Similarity   100.0%;  Pred. No. 1e+02;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KAAAK 5
           |||||
Db      50 KAAAK 54

RESULT 5
Q903W3
ID   Q903W3        PRELIMINARY;       PRT;   60 AA.
AC   Q903W3;
DT   01-MAY-2000 (TREMBLrel. 13, Created)
DT   01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE   HISTONE H1.0.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RA   Jedrusik M.A., Schulze E.;
RT   "The histone H1 complement of Caenorhabditis elegans.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF216291; AAF23175.1; -.
DR   INTERPRO; IPR001386; -.
DR   PRINTS; PR00624; HISTONEH5.
SQ   SEQUENCE   60 AA;  6212 MW;  D01ABB4CBC35566D CRC64;

Query Match             100.0%;  Score 22;  DB 5;  Length 60;
Best Local Similarity   100.0%;  Pred. No. 1e+02;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KAAAK 5
           |||||
Db      44 KAAAK 48

RESULT 6
Q9NFJ7
ID   Q9NFJ7        PRELIMINARY;       PRT;   61 AA.
AC   Q9NFJ7;
DT   01-OCT-2000 (TREMBLrel. 15, Created)
DT   01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE   HISTONE H1.
GN   HIA61.
OS   Trypanosoma brucei gambiense.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=31285;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STIB 755;
RA   Grueter E.;
RL   Thesis (2000), Department of Parasitology, Institute of Zoology,
RL   SWITZERLAND.
DR   EMBL; AJ2B7597; CAB76179.1; -.
SQ   SEQUENCE   61 AA;  6059 MW;  F60DCE6D6D6A73F CRC64;
```

```
Query Match
Best Local Similarity  100.0%;      Score  22;  DB 5;  Length 61;
            Conservative 100.0%;    Pred. No. 1e+02;
Matches    5;  Conservative   0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 KAAAK 5
          |||||
Db     55 KAAAK 59

RESULT 7
Q9NFJ5
ID   Q9NFJ5;        PRELIMINARY;      PRT;    61 AA.
AC   Q9NFJ5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HISTONE H1.
GN   H1A61.
OS   Trypanosoma brucei brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TREU 927/4;
RA   Grueter E.;
RL   Thesis (2000), Department of Parasitology, Institute of Zoology,
RL   SWITZERLAND.
DR   EMBL; AJ287603; CAB76188.1; -.
SQ   SEQUENCE 61 AA; 6012 MW; 7B66DA10A7D33B93 CRC64;

Query Match
Best Local Similarity  100.0%;      Score  22;  DB 5;  Length 61;
            Conservative 100.0%;    Pred. No. 1e+02;
Matches    5;  Conservative   0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 KAAAK 5
          |||||
Db     55 KAAAK 59

RESULT 8
Q17536
ID   Q17536;        PRELIMINARY;      PRT;    62 AA.
AC   Q17536;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   C01B10.5 PROTEIN.
GN   C01B10.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Blanchard M., Bradshaw H.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U58757; AAC47916.1; -.
SQ   SEQUENCE 62 AA; 6453 MW; FF301204D637D406 CRC64;

Query Match
Best Local Similarity  100.0%;      Score  22;  DB 5;  Length 62;
            Conservative 100.0%;    Pred. No. 1.1e+02;
Matches    5;  Conservative   0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 KAAAK 5
          |||||
Db     46 KAAAK 50

RESULT 9
Q9NFJ3
ID   Q9NFJ3;        PRELIMINARY;      PRT;    66 AA.
AC   Q9NFJ3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HISTONE H1.
GN   H1A66.
OS   Trypanosoma brucei brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=427;
RA   Grueter E.;
RL   Thesis (2000), Department of Parasitology, Institute of Zoology,
RL   SWITZERLAND.
DR   EMBL; AJ287608; CAB76193.1; -.
SQ   SEQUENCE 66 AA; 6593 MW; 2D664C3471064DC6 CRC64;

Query Match
Best Local Similarity  100.0%;      Score  22;  DB 5;  Length 66;
            Conservative 100.0%;    Pred. No. 1.1e+02;
Matches    5;  Conservative   0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 KAAAK 5
          |||||
Db     60 KAAAK 64

RESULT 10
Q9NFJ8
ID   Q9NFJ8;        PRELIMINARY;      PRT;    71 AA.
AC   Q9NFJ8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HISTONE H1.
GN   H1A71.
OS   Trypanosoma brucei gambiense.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=31285;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STIB 755;
RA   Grueter E.;
RL   Thesis (2000), Department of Parasitology, Institute of Zoology,
RL   SWITZERLAND.
DR   EMBL; AJ287594; CAB76176.1; -.
SQ   SEQUENCE 71 AA; 7019 MW; 6C7DED4F3B75D8C4 CRC64;

Query Match
Best Local Similarity  100.0%;      Score  22;  DB 5;  Length 71;
            Conservative 100.0%;    Pred. No. 1.2e+02;
Matches    5;  Conservative   0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy      1 KAAAK 5
          |||||
Db     65 KAAAK 69

RESULT 11
Q9N6K5
ID   Q9N6K5      PRELIMINARY;      PRT;      71 AA.
AC   Q9N6K5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HISTONE H1.
GN   H1D71 OR H1B71 OR H1C71.
OS   Trypanosoma brucei brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TREU 927/4;
RA   Grueter E.;
RL   Thesis (2000), Department of Parasitology, Institute of Zoology,
RL   SWITZERLAND.
DR   EMBL; AJ400880; CAB90837.1; -.
DR   EMBL; AJ287602; CAB76187.1; -.
DR   EMBL; AJ287606; CAB76191.1; -.
SQ   SEQUENCE 71 AA; 7019 MW; 406BE51794847C80 CRC64;

Query Match             100.0%;   Score 22;  DB 5;  Length 71;
Best Local Similarity   100.0%;   Pred. No. 1.2e+02;
Matches     5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KAAAK 5
          |||||
Db     65 KAAAK 69

RESULT 12
Q9N6K0
ID   Q9N6K0      PRELIMINARY;      PRT;      71 AA.
AC   Q9N6K0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HISTONE H1.
GN   H1E71 OR H1A71.
OS   Trypanosoma brucei brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TREU 927/4;
RA   Grueter E.;
RL   Thesis (2000), Department of Parasitology, Institute of Zoology,.
DR   EMBL; AJ400881; CAB90839.1; -.
DR   EMBL; AJ287600; CAB76185.1; -.
SQ   SEQUENCE 71 AA; 7005 MW; 478C7B6994847C80 CRC64;

Query Match             100.0%;   Score 22;  DB 5;  Length 71;
Best Local Similarity   100.0%;   Pred. No. 1.2e+02;
Matches     5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KAAAK 5
          |||||
Db     65 KAAAK 69

RESULT 13
P96100
ID   P96100      PRELIMINARY;      PRT;      72 AA.
AC   P96100;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 7.9 KDA PROTEIN.
OS   Thiobacillus ferrooxidans.
OG   Plasmid pTFS.
OC   Bacteria; Proteobacteria; gamma subdivision; Acidithiobacillus.
OX   NCBI_TaxID=920;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC33020;
RX   MEDLINE=98015401; PubMed=9353917;
RA   Dominy C.N., Deane S.M., Rawlings D.E.;
RT   "A geographically widespread plasmid from Thiobacillus ferrooxidans
RT   has genes for ferredoxin-, FNR-, prismane- and NADH-oxidoreductase-
RT   like proteins which are also located on the chromosome.";
RL   Microbiology 143:0-0(0).
DR   EMBL; U73041; AAC80180.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE 72 AA; 7899 MW; E1204B478BAB1AA4 CRC64;

Query Match             100.0%;   Score 22;  DB 2;  Length 72;
Best Local Similarity   100.0%;   Pred. No. 1.2e+02;
Matches     5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KAAAK 5
          |||||
Db     41 KAAAK 45

RESULT 14
Q9NFK0
ID   Q9NFK0      PRELIMINARY;      PRT;      75 AA.
AC   Q9NFK0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HISTONE H1.
GN   H1A75.
OS   Trypanosoma brucei gambiense.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=31285;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STIB 755;
RA   Grueter E.;
RL   Thesis (2000), Department of Parasitology, Institute of Zoology,
RL   SWITZERLAND.
DR   EMBL; AJ287592; CAB76174.1; -.
SQ   SEQUENCE 75 AA; 7597 MW; 130AB85471996052 CRC64;

Query Match             100.0%;   Score 22;  DB 5;  Length 75;
Best Local Similarity   100.0%;   Pred. No. 1.3e+02;
Matches     5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KAAAK 5
          |||||
Db     69 KAAAK 73

RESULT 15
Q9Y5V0
ID   Q9Y5V0      PRELIMINARY;      PRT;      76 AA.
AC   Q9Y5V0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
```

```
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HSPC038 PROTEIN
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ye M., Zhang Q., Zhou J., Shen Y., Guan Z., Wu X., Fan H., Mao H.,
RA   Dai M., Huang Q., Chen S., Chen Z.;
RT   "Human HSPC038 mRNA, complete cds.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF125099; AAD39916.1; -.
DR   INTERPRO; IPR000822; -.
DR   PFAM; PF00096; zf-C2H2; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
KW   Zinc-finger; Metal-binding; DNA-binding.
SEQUENCE   76 AA;  8498 MW;  3EDEABD9FFE529FF CRC64;
```

```
Query Match          100.0%;  Score 22;  DB 4;  Length 75;
Best Local Similarity  100.0%;  Pred. No. 1.3e+02;
Matches      5;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

Qy       1 KAAAK 5
           |||||
Db      30 KAAAK 34
```

```
Search completed: April 24, 2001, 16:40:32
Job time: 422 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001. 16:38:18 ; Search time 115.25 Seconds
   (without alignments)
   2.976 Million cell updates/sec

Title: US-09-340-736-5
Perfect score: 31
Sequence: 1 PGVGVA 6

Scoring table: BLOSUM62
  Gapop 10.0 , Gapext 0.5

tched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
  Maximum Match 100%
  Listing first 45 summaries

Database : A_Geneseq_0401:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	100.0	16	10	P91313	Sequence of beta-t
2	31	100.0	18	13	R26382	Sequence of peptid
3	31	100.0	18	14	R39999	Scintigraph imagin
4	31	100.0	18	17	W11072	Atherosclerotic pl
5	31	100.0	18	18	W31122	Atherosclerotic pl
6	31	100.0	19	12	R15135	Vascular injury af
7	31	100.0	19	14	R42542	Leukocyte-binding
8	31	100.0	19	19	W47337	Inhibitor of LDL b
9	31	100.0	19	20	Y55876	Apolipoprotein fra
10	31	100.0	19	20	Y33101	Elastin derived pe
11	31	100.0	20	14	R40018	Scintigraph imagin

12	31	100.0	20	14	R42535	Leukocyte-binding
13	31	100.0	20	14	R42541	Leukocyte-binding
14	31	100.0	21	14	R42534	Leukocyte-binding
15	31	100.0	22	7	P61339	Sequence of chemot
16	31	100.0	22	12	R14949	Medial middle laye
17	31	100.0	24	14	R42538	Leukocyte-binding
18	31	100.0	28	12	R14979	Part of elastomeri
19	31	100.0	28	12	R14986	Part of elastomeri
20	31	100.0	28	12	R15021	Part of elastomeri
21	31	100.0	28	12	R15028	Part of elastomeri
22	31	100.0	30	12	R14993	Part of elastomeri
23	31	100.0	30	12	R14951	Part of elastomeri
24	31	100.0	30	12	R14958	Part of elastomeri
25	31	100.0	30	12	R15000	Part of elastomeri
26	31	100.0	30	12	R14965	Part of elastomeri
27	31	100.0	30	12	R15014	Part of elastomeri
28	31	100.0	30	12	R14972	Part of elastomeri
29	31	100.0	30	12	R15007	Part of elastomeri
30	31	100.0	42	21	Y74244	Peptide containing
31	31	100.0	112	20	B08629	Human prostate tum
32	31	100.0	114	7	P60728	Synthetic elastome
33	31	100.0	148	20	Y31681	Alanine-containing
34	31	100.0	201	19	W46316	Non-natural polype
35	31	100.0	515	21	Y69135	Amino acid sequenc
36	31	100.0	571	21	Y69071	Amino acid sequenc
37	31	100.0	605	20	Y16791	Truncated plant-op
38	31	100.0	660	20	Y01303	Human tropoelastin
39	31	100.0	698	20	Y01302	Human tropoelastin
40	31	100.0	698	21	Y69059	Amino acid sequenc
41	31	100.0	712	21	B08630	Amino acid sequenc
42	31	100.0	718	19	W61024	Chimeric lepidopte
43	31	100.0	730	19	W46315	Human elastin cont
44	31	100.0	730	21	B08631	Fusion protein com
45	31	100.0	731	21	Y69058	Amino acid sequenc

ALIGNMENTS

RESULT 1
P91313
ID P91313 standard; Protein; 16 AA.
XX
AC P91313;
XX
DT 09-MAR-1992 (first entry)
XX
DE Sequence of beta-turn of a bioelastomeric material.
XX
DB Bioelastomer; elastomeric material; wound dressing; burn;
KW artificial veins; arteries; skin; ligament; biodegradable.
FH
FH Key Location/Qualifiers
FT Misc-difference 1..5
FT   /note= "May be PGVGV,GVGV,VGV,GV,V or a covalent
FT   bond"
FT Misc-difference 12..16
FT   /note= "May be APGVG,APGV,APG,AP,A or a covalent
FT   bond"
FT Misc-difference 6..11
FT   /note= "May present from 2-5,000 times"
FT Misc-difference 1..16
FT   /note= "Must contain at least 18 AAs"
XX
PN WO8910099-A.
XX
PD 02-NOV-1989.
XX
PF 14-APR-1989; 89WO-US01482.
XX
PR 21-APR-1988; 88US-0184407.
XX
PA (UABR-) UAB-RES FOUNDATION.

XX
PI Urry DW;
XX
DR WPI; 1989-339743/46.
XX
PT Elastomeric polypeptide material - a useful for preventing
PT adhesion between tissues and wound repair sites
XX
PS Claim 13; page 89; 93pp; English.
XX
CC The elastomeric material of the invention comprises a bioelastomer
CC contg. repeating elastomeric tetrapeptide or pentapeptide units opt.
CC modified by hexapeptide units, the units consisting of hydrophobic
CC AA residues and Gly residues and existing in a conformation having
CC a beta-turn. The beta-turn of the bioelastomer comprises
CC a polypentapeptide unit of VPGVG (see P91307) and/or IPGVG
CC (see P91308), or a pentapeptide repeating unit of formula XPXXG
CC (see P91311), or a polytetrapeptide of formula VPGG (see P91309),
CC IPGG (see P91310), or XPGG (see P91312); and the hexapeptide
CC repeating unit of formula APGVGV (see P91313).
XX
SQ Sequence 16 AA;

Query Match   100.0%; Score 31; DB 10; Length 16;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
      ||||||
Db 1 pgvgva 6

RESULT 2
ID R26382 standard; peptide; 18 AA.
XX
AC R26382;
XX
DT 25-JAN-1993 (first entry)
XX
DE Sequence of peptides targeted to infections and atherosclerotic
DE plaque.
XX
KW Technetium-99m labelled polypeptide imaging agent;
KW radiolabeled imaging; radiodiagnostic agent.
XX
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Peptide   1..6
FT   /note= "claimed"
XX
PN WO9213572-A.
XX
PD 20-AUG-1992.
XX
PF 07-FEB-1992; 92WO-US00757.
XX
PR 08-FEB-1991; 91US-0653012.
XX
PA (DIAT-) DIATECH INC.
XX
PI Dean RT;
XX
DR WPI; 1992-299767/36.
XX
PT New technetium-99m labelled polypeptide imaging agents - for
PT imaging of clots, tumours, infection sites, atherosclerotic and
PT amyloid plaques or bone, and for visualising organs
XX
PS Claim 6; Page 13; 19pp; English.
XX

CC The binding peptide is covalently linked to a 'Cp(aa)Cp' technetium
CC binding group wherin Cp is a protected cysteine and (aa) is an amino
CC acid. The technetium-99m complexes are used to image target sites
CC within a mammalian body.
XX
SQ Sequence 18 AA;

Query Match   100.0%; Score 31; DB 13; Length 18;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
      ||||||
Db 5 pgvgva 10

RESULT 3
ID R39999 standard; peptide; 18 AA.
XX
AC R39999;
XX
DT 23-MAY-1994 (first entry)
XX
DE Scintigraph imaging agent specific binding peptide.
XX
KW Reagent; site imaging; technetium-99M labelled; peptide.
XX
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Modified-site 18
FT   /note= "C-terminal amide"
XX
PN WO9321962-A.
XX
PD 11-NOV-1993.
XX
PF 19-APR-1993; 93WO-US03687.
XX
PR 30-APR-1992; 92US-0871282.
XX
PA (DIAT-) DIATECH INC.
XX
PI Buttram S, Civitello ER, Dean RT, Lister-James J, McBride W;
XX
DR WPI; 1993-368429/46.
XX
PT Reagents for preparing scintigraphic imaging agents - contg.
PT technetium-99M labelled peptide(s) contg. 3-100 aminoacid(s)
XX
PS Claim 35; Page 39; 55pp; English.
XX
CC The sequence is that of a specific binding peptide used as part of
CC a reagent for preparing a scintigraphic imaging agent for imaging
CC sites within a mammalian body. In this the peptide is covalently
CC linked to a radiolabel-binding moiety which is capable of forming
CC a complex with a radioisotope, pref. technetium-99M.
XX
SQ Sequence 18 AA;

Query Match   100.0%; Score 31; DB 14; Length 18;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
      ||||||
Db 5 pgvgva 10

RESULT 4

W11072
ID   W11072 standard; peptide; 18 AA.
XX
AC   W11072;
XX
DT   03-JUN-1997   (first entry)
XX
DE   Atherosclerotic plaque targetting peptide used in diagnostic imaging.
XX
KW   Leukocyte; target; direct; chelator; radionuclide; radiolabel;
KW   isotope; atherosclerosis; thrombosis; embolism; infection; thrombus;
KW   diagnosis; imaging.
XX
OS   Synthetic.
XX
XX
PN   WO9603427-A1.
XX
PD   09-FEB-1996.
XX
PF   28-APR-1995;   95WO-CA00249.
PR   22-JUL-1994;   94US-0279155.
XX
PA   (RESO-) RESOLUTION PHARM INC.
XX
PI   Goodbody A, Pollak A;
XX
DR   WPI; 1996-116994/12.
XX
PT   New peptide derived radionuclide chelators and metal complexes
PT   useful for diagnostic imaging
XX
PS   Disclosure; Page 7; 30pp; English.
XX
CC   W11072-W11086 are peptides used for targeting agents to an
CC   atherosclerotic plaque or site of infection (no further details are
CC   given in the specification). The peptides may be coupled to a chelator
CC   molecule, which is labelled with a diagnostically useful metal nuclide
CC   to form a peptide derived radionuclide chelator molecules. Such
CC   targetted and labelled chelators are used to detect pathological
CC   conditions by diagnostic imaging Radionuclides used include 99mTc,
CC   64Cu, 67Cu, 97Ru, 105Rh, 109Pd, 186Re, 188Re, 198Au, 199Au, 203Pb, 212Pb
CC   and 212Bi. The coupling of a targeting agent and radionuclide using a
CC   chelating agent is an alternative to the direct labelling of targeting
CC   agents in which radionuclides are typically bound at the more numerous
CC   low-affinity sites, forming unstable complexes. The new conjugates give
CC   better scintigraphic images in rat inflammation studies than known
CC   imaging agents Ga-67, 99mTc-IgG, 111In-WBC and 99mTc-Nanocoll. They image
CC   more rapidly than the known agents and show superior biodistribution.
XX
SQ   Sequence 18 AA;

Query Match   100.0%; Score 31; DB 17; Length 18;
Best Local Similarity   100.0%; Pred. No. 12;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
            ||||||
Db   5 pgvgva 10

RESULT   5
W31122
ID   W31122 standard; peptide; 18 AA.
XX
AC   W31122;
XX
DT   23-JAN-1998   (first entry)
XX
DE   Atherosclerotic plaque- and infection site-targeting peptide.
XX
KW   Target; delivery; radionuclide chelator; diagnosis; therapy;

KW   detection; atherosclerosis; thrombosis; platelet.
XX
OS   Synthetic.
XX
PN   US5659041-A.
XX
PD   19-AUG-1997.
XX
PF   19-JUL-1993;   93US-0092911.
XX
PR   02-SEP-1994;   94US-0299636.
PR   19-JUL-1993;   93US-0092911.
XX
PA   (RESO-) RESOLUTION PHARM INC.
XX
PI   Dunn-Dufault R, Kirby RA, Pollak A;
XX
DR   WPI; 1997-424290/39.
XX
PT   New thio:acetyl-aminoacid hydrazide compounds - useful as chemical
PT   chelator of radionuclides for radio:imaging of target tissues of
PT   diagnostic interest
XX
PS   Disclosure; Column 17-18; 20pp; English.
XX
CC   W31110-W31147 are peptides used for targeting a new hydrazino-type
CC   compound to various sites of disease, e.g. atherosclerotic plaque,
CC   sites of infection, platelets, thrombus or amyloid plaque. The new
CC   compound is a radionuclide chelator and is used to radiolabel the
CC   targeting peptides for the detection and diagnostic imaging of
CC   sites of disease, e.g. amyloid plaques in Alzheimer's disease.
XX
SQ   Sequence 18 AA;

Query Match   100.0%; Score 31; DB 18; Length 18;
Best Local Similarity   100.0%; Pred. No. 12;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
            ||||||
Db   5 pgvgva 10

RESULT   5
R15135
ID   R15135 standard; protein; 19 AA.
XX
AC   R15135;
XX
DT   18-FEB-1992   (first entry)
XX
DE   Vascular injury affinity peptide.
XX
KW   Low density lipoprotein; atherosclerosis.
XX
OS   Synthetic.
XX
PN   WO9116919-A.
XX
PD   14-NOV-1991.
XX
PF   02-MAY-1991;   91WO-US03026.
XX
PR   03-MAY-1990;   90US-0518215.
PR   03-MAY-1990;   90US-0518142.
XX
PA   (NEWE-) NEW ENGLAND DEACON.
XX
PI   Lees RS, Lees AM, Fischman A, Shih IL, Findeis MA;
XX
DR   WPI; 1991-353525/48.
XX

PT Synthetic peptide(s) comprising amphiphilic domain of apoA-I -
PT used to diagnose vascular injury or disease or inhibit binding of
PT low density lipoprotein to vascular walls in treating
PT atherosclerosis
XX
PS Claim 25; Page 48; 66pp; English.
XX
CC The amino acid sequence is that of a synthetic peptide derived from
CC elastin, a vascular associated protein. The peptide, which may opt.
CC be labelled, is used to detect injuries in the vascular system, esp.
CC atherosclerosis in its early stages before it causes stenosis and
CC blood flow disturbances. It can also be used to inhibit binding of
CC low density lipoprotein (LDL) to vascular walls, i.e. to prevent
CC or alleviate atherosclerosis. It is easy to prepare on a large scale
CC and allows vascular regions to be located non-invasively without
CC complex equipment or highly skilled personnel. See also R15126-R15140.
XX
SQ Sequence 19 AA;

Query Match 100.0%; Score 31; DB 12; Length 19;
est Local Similarity 100.0%; Pred. No. 13;
tches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
     |||||
Db 6 pgvgva 11

RESULT 7
ID R42542 standard; Peptide; 19 AA.
XX
AC R42542;
XX
DT 05-AUG-1994 (first entry)
XX
DE Leukocyte-binding peptide which can bind to technetium-99m.
XX
KW Leukocyte binding peptide; elastin; scintigraphic imaging;
KW inflammation site; technetium 99m.
XX
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Modified-site   1
FT   /label= OTHER.
FT   /note= "BAT, i.e. N6,N9-bis(2-methyl-
FT   2-mercaptopropyl)-6,9-diazaoananoic acid"
FT Modified-site   19
FT   /note= "Gly-NH2"
XX
PN WO9317719-A.
XX
PD 16-SEP-1993.
XX
PF 12-MAR-1993; 93WO-US02320.
XX
PR 13-MAR-1992; 92US-0851074.
XX
PA (DIAT-) DIATECH INC.
XX
PI Buttram S, Dean RT, Lees RS, Lister-James J;
XX
DR WPI; 1993-303154/38.
XX
PT Scintigraphic imaging agent for sites of inflammation -
PT comprising leukocyte-binding peptide bound technetium-99m via
PT binding moiety
XX
PS Claim 14; Page 31; 40pp; English.
XX
CC New peptides are claimed which are leukocyte binding peptides

CC having covalently bound to them a moiety which can bind a Tc-99m
CC radiolabel. The peptides having Tc-99m bound to them are useful as
CC scintigraphic imaging agents for imaging sites of infection and
CC inflammation in the mammalian body, e.g. caused by ischaemia,
CC inflammatory bowel disorder, arthritis or tumours.
CC The present sequence is a specifically claimed example of
CC such a peptide derived from elastin.
XX
SQ Sequence 19 AA;

Query Match 100.0%; Score 31; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
     |||||
Db 6 pgvgva 11

RESULT 8
ID W47337 standard; peptide; 19 AA.
XX
AC W47337;
XX
DT 01-JUN-1998 (first entry)
XX
DE Inhibitor of LDL binding to vascular wall.
XX
KW Binding inhibitor; low-density lipoprotein; LDL; vascular wall;
KW vascular injury; elastin; collagen; prevention; treatment;
KW vascular disease; atherosclerosis.
XX
OS Synthetic.
XX
PN US5726153-A.
XX
PD 10-MAR-1998.
XX
PF 06-JUN-1995; 95US-0468543.
XX
PR 06-JUN-1995; 95US-0469692.
PR 02-MAY-1988; 88US-0189130.
PR 03-MAY-1990; 90US-0518142.
PR 03-MAY-1990; 90US-0518215.
PR 02-MAY-1990; 90US-0694929.
PR 16-APR-1991; 91US-0694929.
PR 24-FEB-1993; 93US-0048569.
PR 28-FEB-1994; 94US-0201057.
PR 28-FEB-1995; 95US-0398046.
PR 06-JUN-1995; 95US-0468543.
XX
PA (NEWE-) NEW ENGLAND DEACONESS HOSPITAL.
XX
PI Findeis MA, Fischman A, Lees AM, Lees RS, Shih I;
XX
DR WPI; 1998-192802/17.
XX
PT Inhibiting binding of low-density lipoprotein to vascular walls - by
PT administering peptide with affinity for vascular injury sites
XX
PS Disclosure; Column 5; 31pp; English.
XX
CC A novel method for inhibiting the binding of a low-density
CC lipoprotein (LDL) to vascular walls in vivo, comprises
CC administering a synthetic water soluble peptide containing an
CC amphiphilic domain and having affinity for sites of vascular
CC injury, e.g. the present peptide.
CC The peptide inhibits the binding of LDL to vascular wall
CC components, e.g. elastin and collagen, and so can be used to
CC prevent or treat vascular diseases, e.g. atherosclerosis.
XX
SQ Sequence 19 AA;

Query Match   100.0%;   Score 31;   DB 19;   Length 19;
Best Local Similarity   100.0%;   Pred. No. 13;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 PGVGVA 6
        ||||||
Db   6 pgvgva 11

RESULT   9
Y55876
ID   Y55876 standard; peptide; 19 AA.
XX
AC   Y55876;
XX
PD   01-FEB-2000 (first entry)
XX
    Apolipoprotein fragment peptide #26 for vascular disease imaging.
XX
KW   Diagnosis; water soluble; amphiphilic domain; affinity; vascular injury;
KW   detection; imaging; disease: atherosclerosis; apolipoprotein.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   US5972830-A.
XX
PD   26-OCT-1999.
XX
PF   28-FEB-1995;   95US-0398046.
XX
PR   02-MAY-1991;   91US-0694929.
PR   16-APR-1993;   93US-0048569.
PR   24-FEB-1994;   94US-0201057.
PR   02-MAY-1988;   88US-0189130.
PR   03-MAY-1990;   90US-0518142.
PR   03-MAY-1990;   90US-0518215.
XX
PA   (NEWE-) NEW ENGLAND DEACONESS HOSPITAL.
XX
PI   Lees AM, Fischman A, Shih I, Findeis MA, Lees RS;
XX
DR   WPI; 1999-632641/54.
XX
    New diagnostic synthetic peptides which have affinity for and
    accumulate at a site of vascular injury useful for detection and
    imaging of vascular disease such as atherosclerosis -
XX
PS   Disclosure: Column 5; 30pp; English.
XX
CC   The peptides Y55851-Y55889 represent examples of diagnostic, synthetic
CC   peptides which carry a detectable label, contain 30 or fewer amino acids,
CC   are water soluble, contain an amphiphilic domain and have affinity for,
CC   and propensity to accumulate at, a site of vascular injury. They are
CC   preferably derived from the amino acid sequence of apolipoprotein. The
CC   peptides can be used for the detection or imaging of a vascular injury
CC   or disease, e.g. atherosclerosis.
XX
SQ   19 AA;

Query Match   100.0%;   Score 31;   DB 20;   Length 19;
Best Local Similarity   100.0%;   Pred. No. 13;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 PGVGVA 6
        ||||||
Db   6 pgvgva 11

RESULT   10

Y33101
ID   Y33101 standard; peptide; 19 AA.
XX
AC   Y33101;
XX
DT   15-NOV-1999 (first entry)
XX
DE   Elastin derived peptide 4.
XX
KW   Alipoprotein B; alipoprotein AI; elastin; vascular imaging; detection;
KW   injury; vascular system: atherosclerosis; re-endotheliizing; noninvasive;
KW   aortic lesion; trauma; lipoprotein accumulation.
XX
OS   Synthetic.
XX
PN   US5955055-A.
XX
PD   21-SEP-1999.
XX
PF   06-JUN-1995;   95US-0469692.
XX
PR   02-MAY-1991;   91US-0694929.
PR   02-MAY-1988;   88US-0189130.
PR   03-MAY-1990;   90US-0518142.
PR   03-MAY-1990;   90US-0518215.
PR   16-APR-1993;   93US-0048569.
PR   24-FEB-1994;   94US-0201057.
PR   28-FEB-1995;   95US-0398046.
PR   06-JUN-1995;   95US-0469692.
XX
PA   (NEWE-) NEW ENGLAND DEACONESS HOSPITAL.
XX
PI   Findeis MA, Fischman A, Lees AM, Lees RS, Shih I;
XX
DR   WPI; 1999-539543/45.
XX
PT   Detecting vascular injuries using a labeled peptide useful for the
PT   diagnosis and monitoring of atherosclerosis
XX
PS   Disclosure; Column 27-28; 31pp; English.
XX
CC   This invention describes a novel method (I) for detecting injuries in a
CC   vascular system using a labeled synthetic peptide with an affinity for,
CC   and propensity to accumulate at, a site of vascular injury. The method is
CC   particularly suitable for detecting and monitoring atherosclerosis. It
CC   has been found that the synthetic peptide accumulates at the healing
CC   (re-endothelizing) edge of aortic lesions produced by trauma. These
CC   lesions resemble human arteriosclerosis in many important respects
CC   (including accumulation of lipoproteins and other pathological changes),
CC   The ability of the synthetic peptides to localize at the trauma site, and
CC   to permit imaging, may be used to visualize vascular disease. The method
CC   is noninvasive and the peptides used to target vascular lesions may be
CC   produced inexpensively, quickly and in large quantities. Y33088-Y33114
CC   represent synthetic peptides derived from alipoprotein B, alipoprotein AI
CC   and elastin which are used in the method of the invention.
XX
SQ   19 AA;

Query Match   100.0%;   Score 31;   DB 20;   Length 19;
Best Local Similarity   100.0%;   Pred. No. 13;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 PGVGVA 6
        ||||||
Db   6 pgvgva 11

RESULT   11
R40018
ID   R40018 standard; peptide; 20 AA.
XX
AC   R40018;

```
XX
DT   23-MAY-1994  (first entry)
XX
DE   Scintigraph imaging agent specific binding peptide.
XX
KW   Reagent; site imaging; technetium-99M labelled; peptide.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   20
FT                   /note= "C-terminal amide"
XX
PN   WO9321962-A.
XX
PD   11-NOV-1993.
XX
PF   19-APR-1993;  93WO-US03687.
XX
PR   30-APR-1992;  92US-0871282.
XX
PA   (DIAT-) DIATECH INC.
XX
PI   Buttram S,  Civitello ER,  Dean RT,  Lister-James J,  McBride W;
XX
DR   WPI: 1993-368429/46.
XX
PT   Reagents for preparing scintigraphic imaging agents - contg.
PT   technetium-99M labelled peptide(s) contg. 3-100 aminoacid(s)
XX
PS   Claim 35; Page 39; 55pp; English.
XX
CC   The sequence is that of a specific binding peptide used as part of
CC   a reagent for preparing a scintigraphic imaging agent for imaging
CC   sites within a mammalian body. In this the peptide is covalently
CC   linked to a radiolabel-binding moiety which is capable of forming
CC   a complex with a radioisotope, pref. technetium-99M.
XX
SQ   Sequence  20 AA;

Query Match       100.0%;  Score 31;  DB 14;  Length 20;
Best Local Similarity  100.0%;  Pred. No. 14;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   1 PGVGVA 6
       ||||||
Db   7 pgvgva 12


RESULT 12
ID   R42535 standard; Peptide; 20 AA.
XX
AC   R42535;
XX
DT   05-AUG-1994  (first entry)
XX
DE   Leukocyte-binding peptide which can bind to technetium-99M.
XX
KW   Leukocyte binding peptide; elastin; scintigraphic imaging;
KW   inflammation site; technetium 99m.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "picolinoyl-Gly"
FT   Modified-site   2
FT                   /note= "acetamidomethyl-Cys"
FT   Modified-site   20
FT                   /note= "Gly-NH2"
XX
```

```
PN   WO9317719-A.
XX
PD   16-SEP-1993.
XX
PF   12-MAR-1993;  93WO-US02320.
XX
PR   13-MAR-1992;  92US-0851074.
XX
PA   (DIAT-) DIATECH INC.
XX
PI   Buttram S,  Dean RT,  Lees RS,  Lister-James J;
XX
DR   WPI: 1993-303154/38.
XX
PT   Scintigraphic imaging agent for sites of inflammation -
PT   comprising leukocyte-binding peptide bound technetium-99m via
PT   binding moiety
XX
PS   Claim 14; Page 31; 40pp; English.
XX
CC   New peptides are claimed which are leukocyte binding peptides
CC   having covalently bound to them a moiety which can bind a Tc-9m
CC   radiolabel. The peptides having Tc-99m bound to them are useful as
CC   scintigraphic imaging agents for imaging sites of infection and
CC   inflammation in the mammalian body. e.g. caused by ischaemia,
CC   inflammatory bowel disorder, arthritis or tumours.
CC   The present sequence is a specifically claimed example of
CC   such a peptide derived from elastin.
XX
SQ   Sequence  20 AA;

Query Match       100.0%;  Score 31;  DB 14;  Length 20;
Best Local Similarity  100.0%;  Pred. No. 14;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   1 PGVGVA 6
       ||||||
Db   7 pgvgva 12


RESULT 13
R42541
ID   R42541 standard; Peptide; 20 AA.
XX
AC   R42541;
XX
DT   05-AUG-1994  (first entry)
XX
DE   Leukocyte-binding peptide which can bind to technetium-99m.
XX
KW   Leukocyte binding peptide; elastin; scintigraphic imaging;
KW   inflammation site; technetium 99m.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1
FT                   /note= "picolinoyl-Gly"
FT   Modified-site   20
FT                   /note= "Gly-NH2"
XX
PN   WO9317719-A.
XX
PD   16-SEP-1993.
XX
PF   12-MAR-1993;  93WO-US02320.
XX
PR   13-MAR-1992;  92US-0851074.
XX
PA   (DIAT-) DIATECH INC.
XX
PI   Buttram S,  Dean RT,  Lees RS,  Lister-James J;
XX
```

```
XX
DR   WPI; 1993-303154/38.
XX
PT   Scintigraphic imaging agent for sites of inflammation -
PT   comprising leukocyte-binding peptide bound technetium-99m via
PT   binding moiety
XX
PS   Claim 14; Page 31; 40pp; English.
XX
CC   New peptides are claimed which are leukocyte binding peptides
CC   having covalently bound to them a moiety which can bind a Tc-99m
CC   radiolabel. The peptides having Tc-99m bound to them are useful as
CC   scintigraphic imaging agents for imaging sites of infection and
CC   inflammation in the mammalian body, e.g. caused by ischaemia,
CC   inflammatory bowel disorder, arthritis or tumours.
CC   The present sequence is a specifically claimed example of
CC   such a peptide derived from elastin.
XX
SQ   Sequence   20 AA;

Query Match          100.0%;   Score 31;   DB 14;   Length 20;
Best Local Similarity 100.0%;   Pred. No. 14;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 PGVGVA 6
       ||||||
Db   7 pgvgva 12


RESULT 14
R42534
ID   R42534 standard; Peptide; 21 AA.
XX
AC   R42534;
XX
DT   05-AUG-1994  (first entry)
XX
DE   Leukocyte-binding peptide which can bind to technetium-99m.
XX
KW   Leukocyte binding peptide; elastin; scintigraphic imaging;
KW   inflammation site; technetium 99m.
XX
OS   Synthetic.
XX
XX
Key            Location/Qualifiers
FT   Modified-site   1
FT                   /note= "acetamidomethyl-Cys"
FT   Modified-site   3
FT                   /note= "acetamidomethyl-Cys"
FT   Modified-site   21
FT                   /note= "Gly-NH2"
XX
PN   WO9317719-A.
XX
PD   16-SEP-1993.
XX
PF   12-MAR-1993;   93WO-US02320.
XX
PR   13-MAR-1992;   92US-0851074.
XX
PA   (DIAT-) DIATECH INC.
XX
PI   Buttram S,  Dean RT,  Lees RS,  Lister-James J;
XX
DR   WPI; 1993-303154/38.
XX
PT   Scintigraphic imaging agent for sites of inflammation -
PT   comprising leukocyte-binding peptide bound technetium-99m via
PT   binding moiety
XX
PS   Claim 14; Page 31; 40pp; English.
XX
```

```
CC   New peptides are claimed which are leukocyte binding peptides
CC   having covalently bound to them a moiety which can bind a Tc-99m
CC   radiolabel. The peptides having Tc-99m bound to them are useful as
CC   scintigraphic imaging agents for imaging sites of infection and
CC   inflammation in the mammalian body, e.g. caused by ischaemia,
CC   inflammatory bowel disorder, arthritis or tumours.
CC   The present sequence is a specifically claimed example of
CC   such a peptide derived from elastin.
XX
SQ   Sequence   21 AA;

Query Match          100.0%;   Score 31;   DB 14;   Length 21;
Best Local Similarity 100.0%;   Pred. No. 14;
Matches   6;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 PGVGVA 6
       ||||||
Db   8 pgvgva 13


RESULT 15
P61339
ID   P61339 standard; peptide; 22 AA.
XX
AC   P61339;
XX
DT   30-JUL-1991  (first entry)
XX
DE   Sequence of chemotactic peptide which includes a repeat hexapeptide
DE   sequence found in tropoelastin from vascular wall.
XX
KW   Elastic fibre-forming fibroblasts; prosthetic device.
XX
FH   Key            Location/Qualifiers
FT   Misc-difference 1..4
FT                   /note= "residues 1-4,1-3,1-2 or 1 may be absent;
FT                   N-terminal is bonded to H or biocompatible
FT                   N-terminal gp"
FT   Misc-difference 19..22
FT                   /note= "residues 19-22,20-22;21-22 or 22 may be
FT                   absent; C-terminal is bonded to OH.OB3 or
FT                   biocompatible C-terminal gp; B3- non-toxic
FT                   metal ion"
FT   Misc-difference 12..17
FT                   /note= "repeat hexapeptide; present 1-100 times"
XX
PN   US4605413-A.
XX
PD   12-AUG-1986.
XX
PF   19-SEP-1983;   83US-0533670.
XX
PR   19-SEP-1983;   83US-0533670.
XX
PA   (UYAL-) UNIV ALABAMA.
XX
PI   Urry DW,  Senior RM;
XX
DR   WPI; 1986-232093/35.
XX
PT   Prosthetic device, e.g. artificial blood vessel or skin - having
PT   chemo-tactic peptide in its surface to enhance invasion of
PT   elastic fibre-forming fibroblasts
XX
PS   Claim 1; column 8; 10pp; English.
XX
CC   The repeat sequence is chemotactic for fibroblasts which synthesise
CC   elastic fibre precursor protein in biological systems.
XX
SQ   Sequence   22 AA;
```

Query Match   100.0%; Score 31; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 15;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
        ||||||
Db   1 pgvgva 6

Search completed: April 24, 2001. 16:38:19
Job time: 419 sec

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   April 24, 2001, 16:36:23 ; Search time 62.39 Seconds
                                     (without alignments)
                                     1.847 Million cell updates/sec

Title:          US-09-340-736-5
Perfect score:  31
Sequence:       1 PGVGVA 6

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Matched:        185757 segs, 19210857 residues

Total number of hits satisfying chosen parameters:   185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      Issued_Patents_AA:*
                1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	100.0	6	2	US-08-911-364-5	Sequence 5, Appli
2	31	100.0	18	1	US-08-127-351-34	Sequence 34, Appl
3	31	100.0	18	1	US-08-480-367B-34	Sequence 34, Appl
4	31	100.0	18	1	US-08-487-221A-34	Sequence 34, Appl
5	31	100.0	18	1	US-08-480-370-34	Sequence 34, Appl
6	31	100.0	18	1	US-08-279-636-13	Sequence 13, Appl
7	31	100.0	18	1	US-08-279-155-14	Sequence 14, Appl
8	31	100.0	18	1	US-08-464-456-13	Sequence 13, Appl
9	31	100.0	18	1	US-08-486-135-1	Sequence 1, Appl
10	31	100.0	18	1	US-08-703-988A-14	Sequence 14, Appl
11	31	100.0	18	1	US-08-463-052-13	Sequence 13, Appl
12	31	100.0	18	2	US-08-480-551-1	Sequence 1, Appl
13	31	100.0	18	2	US-08-612-842-14	Sequence 14, Appl
14	31	100.0	18	2	US-08-290-853-28	Sequence 28, Appl
15	31	100.0	18	1	US-08-468-543-14	Sequence 14, Appl
16	31	100.0	19	1	US-08-469-692-14	Sequence 14, Appl
17	31	100.0	19	2	US-08-398-046-14	Sequence 14, Appl
18	31	100.0	20	1	US-08-486-135-19	Sequence 19, Appl
19	31	100.0	20	1	US-08-470-152-19	Sequence 19, Appl
20	31	100.0	21	1	US-08-472-535-11	Sequence 11, Appl
21	31	100.0	21	1	US-08-472-535-12	Sequence 12, Appl
22	31	100.0	21	1	US-08-484-774-11	Sequence 11, Appl
23	31	100.0	21	1	US-08-290-853-2	Sequence 2, Appl
24	31	100.0	21	1	US-08-290-853-7	Sequence 7, Appl
25	31	100.0	21	2	US-08-290-853-7	Sequence 7, Appl
26	31	100.0	21	1	US-08-290-853-2	Sequence 2, Appl
27	31	100.0	21	2	US-08-290-853-17	Sequence 17, Appl
28	31	100.0	21	3	US-08-266-178A-11	Sequence 11, Appl
29	31	100.0	21	3	US-08-266-178A-12	Sequence 12, Appl
30	31	100.0	21	3	US-08-472-535-15	Sequence 15, Appl
31	31	100.0	24	1	US-08-484-774-15	Sequence 15, Appl
32	31	100.0	24	2	US-08-290-853-13	Sequence 13, Appl
33	31	100.0	24	3	US-08-266-178A-15	Sequence 15, Appl
34	31	100.0	27	6	5250516-15	Patent No. 5250516
35	31	100.0	201	2	US-08-911-364-2	Sequence 2, Appli
36	31	100.0	718	2	US-08-731-079A-1	Sequence 1, Appli
37	31	100.0	731	2	US-08-911-364-1	Sequence 1, Appli
38	31	100.0	792	2	US-08-678-039A-40	Sequence 40, Appl
39	31	100.0	1148	2	US-08-349-867-23	Sequence 23, Appl
40	31	100.0	1148	1	US-08-349-867-27	Sequence 27, Appl
41	31	100.0	1148	1	US-08-239-476-23	Sequence 23, Appl
42	31	100.0	1148	1	US-08-239-476-27	Sequence 27, Appl
43	31	100.0	1148	2	US-08-598-305A-23	Sequence 23, Appl
44	31	100.0	1148	2	US-08-598-305A-27	Sequence 27, Appl
45	31	100.0	1148	2	US-08-598-305A-35	Sequence 35, Appl

ALIGNMENTS

```
RESULT 1
US-08-911-364-5
; Sequence 5, Application US/08911364
; Patent No. 5969106
; GENERAL INFORMATION:
;  APPLICANT: ROTHSTEIN, Aser
;  APPLICANT: KEELY, Fred W.
;  APPLICANT: ROTHSTEIN, Steven J.
;  TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
;  TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
;  NUMBER OF SEQUENCES: 8
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FOLEY & LARDNER
;   STREET: 3000 K Street, N.W.
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20007-5109
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/911,364
;   FILING DATE: 07-AUG-1997
;   CLASSIFICATION: 530
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/023,552
;   FILING DATE: 07-AUG-1996
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Bent, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 041082/0104
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 672-5300
;   TELEFAX: (202) 672-5399
;  INFORMATION FOR SEQ ID NO: 5:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 6 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;  MOLECULE TYPE: peptide
US-08-911-364-5

  Query Match             100.0%;  Score 31;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

Qy   1 PGVGVA 6
         ||||||
Db   1 PGVGVA 6

RESULT 2
US-08-127-351-34
; Sequence 34, Application US/08127351
; Patent No. 5449761
; GENERAL INFORMATION:
;   APPLICANT: BELINKA Jr, BENJAMIN A.
;   APPLICANT: COUGHLIN, DANIEL J.
;   APPLICANT: ALVAREZ, VERNON L.
;   APPLICANT: WOOD, RICHARD
;   TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE
;   TITLE OF INVENTION: CONSTRUCTS
;   NUMBER OF SEQUENCES: 56
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER &
;   ADDRESSEE: NEUSTADT,
;   ADDRESSEE: P.C.
;   STREET: 1755 S. Jefferson Davis Highway, Suite 400
;   CITY: Arlington
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 22202
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/127,351
;   FILING DATE: 28-SEP-1993
;   CLASSIFICATION: 534
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Villacorta, Gilberto M.
;   REGISTRATION NUMBER: 34,038
;   REFERENCE/DOCKET NUMBER: 4980-004-44
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 413-3000
;   TELEFAX: (703) 413-2220
;   TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 34:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 18 amino acids
;   TYPE: amino acid
;   TOPOLOGY: unknown
;   MOLECULE TYPE: peptide
US-08-127-351-34

Query Match   100.0%; Score 31; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
         ||||||
Db   5 PGVGVA 10

RESULT 3
US-08-480-367B-34
; Sequence 34, Application US/08480367B
; Patent No. 5578288
; GENERAL INFORMATION:
;   APPLICANT: BELINKA Jr, BENJAMIN A.
;   APPLICANT: COUGHLIN, DANIEL J.
;   APPLICANT: ALVAREZ, VERNON L.
;   APPLICANT: WOOD, RICHARD
;   TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE
;   TITLE OF INVENTION: CONSTRUCTS

;   NUMBER OF SEQUENCES: 56
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: LOWE, PRICE, LeBLANC & BECKER
;   STREET: 99 Canal Center Plaza, Suite 300
;   CITY: Alexandria
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 22314
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/480,367B
;   FILING DATE: 07-06-95
;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Villacorta, Gilberto M.
;   REGISTRATION NUMBER: 34,038
;   REFERENCE/DOCKET NUMBER: 2654-002A
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 684-1111
;   TELEFAX: (703) 684-1124
;   TELEX:
; INFORMATION FOR SEQ ID NO: 34:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 18 amino acids
;   TYPE: amino acid
;   TOPOLOGY: unknown
;   MOLECULE TYPE: peptide
US-08-480-367B-34

Query Match   100.0%; Score 31; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
         ||||||
Db   5 PGVGVA 10

RESULT 4
US-08-487-221A-34
; Sequence 34, Application US/08487221A
; Patent No. 5593556
; GENERAL INFORMATION:
;   APPLICANT: BELINKA Jr, BENJAMIN A.
;   APPLICANT: COUGHLIN, DANIEL J.
;   APPLICANT: ALVAREZ, VERNON L.
;   APPLICANT: WOOD, RICHARD
;   TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE
;   TITLE OF INVENTION: CONSTRUCTS
;   NUMBER OF SEQUENCES: 56
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER &
;   ADDRESSEE: NEUSTADT, P.C.
;   STREET: 1755 S. Jefferson Davis Highway, Suite 400
;   CITY: Arlington
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 22202
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/487,221A
;   FILING DATE: 07-JUN-1995
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:

```
;     TOPOLOGY: unknown
;     MOLECULE TYPE: peptide
US-08-480-370-34

Query Match           100.0%;   Score 31;  DB 1;  Length 18;
Best Local Similarity 100.0%;   Pred. No. 4.4;
Matches    6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 PGVGVA 6
          ||||||
Db     5 PGVGVA 10

RESULT 6
US-08-299-636-13
; Sequence 13, Application US/08299636
; Patent No. 5659041
; GENERAL INFORMATION:
;   APPLICANT: POLLAK, Alfred
;   APPLICANT: KIRBY, Robert A.
;   APPLICANT: DUNN-DUFAULT, Robert
;   TITLE OF INVENTION: HYDRAZINO-TYPE RADIONUCLIDE CHELATORS
;   TITLE OF INVENTION: HAVING AN N3S CONFIGURATION
;   NUMBER OF SEQUENCES: 39
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Foley & Lardner
;     STREET: 3000 K Street, N.W., Suite 500
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: USA
;     ZIP: 20007-5109
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/299,636
;     FILING DATE: 02-SEP-1994
;     CLASSIFICATION: 534
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/092,911
;     FILING DATE: 18-JUL-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BENT, Stephen A.
;     REGISTRATION NUMBER: 29,768
;     REFERENCE/DOCKET NUMBER: 16777/262/ALLE
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (202)672-5300
;     TELEFAX: (202)672-5399
;     TELEX: 904136
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 18 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
US-08-299-636-13

Query Match           100.0%;   Score 31;  DB 1;  Length 18;
Best Local Similarity 100.0%;   Pred. No. 4.4;
Matches    6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 PGVGVA 6
          ||||||
Db     5 PGVGVA 10

RESULT 7
US-08-279-155-14
; Sequence 14, Application US/08279155
```

```
;     APPLICATION NUMBER: US 08/127,351
;     FILING DATE: 28-SEP-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Villacorta, Gilberto M.
;     REGISTRATION NUMBER: 34,038
;     REFERENCE/DOCKET NUMBER: 4980-004-44
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703) 413-3000
;     TELEFAX: (703) 413-2220
;     TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 34:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 18 amino acids
;     TYPE: amino acid
;     TOPOLOGY: unknown
;     MOLECULE TYPE: peptide
US-08-487-221A-34

Query Match           100.0%;   Score 31;  DB 1;  Length 18;
Best Local Similarity 100.0%;   Pred. No. 4.4;
Matches    6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 PGVGVA 6
          ||||||
Db     5 PGVGVA 10

RESULT 5
US-08-480-370-34
; Sequence 34, Application US/08480370
; Patent No. 5609847
; GENERAL INFORMATION:
;   APPLICANT: BELINKA Jr, BENJAMIN A.
;   APPLICANT: COUGHLIN, DANIEL J.
;   APPLICANT: ALVAREZ, VERNON L.
;   APPLICANT: WOOD, RICHARD
;   TITLE OF INVENTION: METAL-BINDING TARGETED POLYPEPTIDE
;   TITLE OF INVENTION: CONSTRUCTS
;   NUMBER OF SEQUENCES: 56
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER &
;     ADDRESSEE: NEUSTADT, P.C.
;     STREET: 1755 S. Jefferson Davis Highway, Suite 400
;     CITY: Arlington
;     STATE: Virginia
;     COUNTRY: U.S.A.
;     ZIP: 22202
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/480,370
;     FILING DATE:
;   PRIOR APPLICATION DATA:
;     CLASSIFICATION: 424
;     APPLICATION NUMBER: US 08/127,351
;     FILING DATE: 28-SEP-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Villacorta, Gilberto M.
;     REGISTRATION NUMBER: 34,038
;     REFERENCE/DOCKET NUMBER: 4980-004-44
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703) 413-3000
;     TELEFAX: (703) 413-2220
;     TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 34:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 18 amino acids
;     TYPE: amino acid
```

Patent No. 5662885
GENERAL INFORMATION:
; APPLICANT: POLLAK, Alfred
; APPLICANT: GOODBODY, Anne
; TITLE OF INVENTION: PEPTIDE DERIVED RADIONUCLIDE CHELATORS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAIDO, MARMELSTEIN, MURRAY & ORAM LLP
; STREET: 655 Fifteenth Street, N. W., Suite 330 - G
; STREET: Street Lobby
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/279,155
; FILING DATE: 22-JUL-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: MURRAY, Robert B.
; REGISTRATION NUMBER: 22,980
; REFERENCE/DOCKET NUMBER: P8074-4005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-279-155-14

Query Match   100.0%; Score 31; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
      ||||||
Db 5 PGVGVA 10

RESULT 8
08-464-456-13
;quence 13, Application US/08464456
;Patent No. 5681541
;GENERAL INFORMATION:
; APPLICANT: Dean, Richard T
; TITLE OF INVENTION: Technetium-99m Labeled Polypeptides for
; TITLE OF INVENTION: Imaging
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, Ltd.
; STREET: 10 South Wacker Drive Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,456
; FILING DATE: 05-JUN-1995

; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5681541nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 90,1104-V
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 715 1000
; TELEFAX: 312 715 1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-464-456-13

Query Match   100.0%; Score 31; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
      ||||||
Db 5 PGVGVA 10

RESULT 9
US-08-486-135-1
; Sequence 1, Application US/08486135
; Patent No. 5720934
; GENERAL INFORMATION:
; APPLICANT: Dean, Richard T
; APPLICANT: Buttram, Scott
; APPLICANT: McBride, William
; APPLICANT: Lister-James, John
; APPLICANT: Civitello, Edgar R
; TITLE OF INVENTION: Technetium-99m Labeled Peptides for
; TITLE OF INVENTION: Imaging
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,135
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5720934nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 92,205-N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 23..24

;   OTHER INFORMATION: /label= AMIDE
;   OTHER INFORMATION: /note= "The carboxyl terminus is an amide."
US-08-486-135-1

Query Match   100.0%; Score 31; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
      |||||
Db   5 PGVGVA 10

RESULT 10
US-08-703-988A-14
; Sequence 14, Application US/08703988A
; Patent No. 5780006
; GENERAL INFORMATION:
;   APPLICANT: POLLAK, Alfred
;   APPLICANT: GOODBODY, Anne
;   TITLE OF INVENTION: PEPTIDE DERIVED RADIONUCLIDE
;   TITLE OF INVENTION: CHELATORS
;   NUMBER OF SEQUENCES: 38
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: NIKAIDO, MARMELSTEIN, MURRAY & ORAM
;   ADDRESSEE: LLP
;   STREET: 655 Fifteenth Street, N. W., Suite 330
;   STREET: G Street Lobby
;   CITY: Washington
;   STATE: DC
;   COUNTRY: USA
;   ZIP: 20005-5701
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/703,988A
;   FILING DATE: 28-AUG-1996
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/279,155
;   FILING DATE: 22-JUL-1994
;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: MURRAY, Robert B.
;   REGISTRATION NUMBER: 22,980
;   REFERENCE/DOCKET NUMBER: P8074-6011
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202/638-5000
;   TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 18 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-703-988A-14

Query Match   100.0%; Score 31; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
      |||||
Db   5 PGVGVA 10

RESULT 11
US-08-470-152-1
; Sequence 1, Application US/08470152
; Patent No. 5780007
; GENERAL INFORMATION:
;   APPLICANT: Dean, Richard T
;   APPLICANT: Buttram, Scott
;   APPLICANT: McBride, William
;   APPLICANT: Lister-James, John
;   APPLICANT: Civitello, Edgar R
;   TITLE OF INVENTION: Technetium-99m Labeled Peptides for
;   TITLE OF INVENTION: Imaging
;   NUMBER OF SEQUENCES: 19
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Banner & Allegretti, Ltd.
;   STREET: 10 South Wacker Drive, Suite 3000
;   CITY: Chicago
;   STATE: IL
;   COUNTRY: USA
;   ZIP: 60606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/470,152
;   FILING DATE: 06-JUN-1995
;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: No. 5780007nan., Kevin E
;   REGISTRATION NUMBER: 35,303
;   REFERENCE/DOCKET NUMBER: 92,205-L
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312-715-1000
;   TELEFAX: 312-715-1234
;   TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 18 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   FEATURE:
;   NAME/KEY: Modified-site
;   LOCATION: 23..24
;   OTHER INFORMATION: /label= AMIDE
;   OTHER INFORMATION: /note= "The carboxyl terminus is an amide."
US-08-470-152-1

Query Match   100.0%; Score 31; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
      |||||
Db   5 PGVGVA 10

RESULT 12
US-08-463-052-13
; Sequence 13, Application US/08463052
; Patent No. 5788960
; GENERAL INFORMATION:
;   APPLICANT: Dean, Richard T
;   TITLE OF INVENTION: Technetium-99m Labeled Polypeptides for
;   TITLE OF INVENTION: Imaging
;   NUMBER OF SEQUENCES: 37
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Banner & Allegretti, Ltd.
;   STREET: 10 South Wacker Drive Suite 3000
;   CITY: Chicago

```
;      STATE: Illinois
;      COUNTRY: USA
;      ZIP: 60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/463,052
;      FILING DATE: 05-JUN-1995
;      CLASSIFICATION: 530
;    ATTORNEY/AGENT INFORMATION:
;      NAME: No 578896nan, Kevin E
;      REGISTRATION NUMBER: 35,303
;      REFERENCE/DOCKET NUMBER: 90,1104-V
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 312 715 1000
;      TELEFAX: 312 715 1234
;      TELEX: 910-221-5317
;  INFORMATION FOR SEQ ID NO: 13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 18 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
US-08-463-052-13

Query Match             100.0%;  Score 31;  DB 1;  Length 18;
Best Local Similarity   100.0%;  Pred. No. 4.4;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PGVGVA 6
              ||||||
Db          5 PGVGVA 10

RESULT 13
US-08-480-551-13
; Sequence 13, Application US/08480551
; Patent No. 5811394
; GENERAL INFORMATION:
;    APPLICANT: Dean, Richard T
;    TITLE OF INVENTION: Technetium-99m Labeled Polypeptides for
;    TITLE OF INVENTION: Imaging
;    NUMBER OF SEQUENCES: 37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Allegretti & Witcoff, Ltd.
;      STREET: 10 South Wacker Drive Suite 3000
;      CITY: Chicago
;      STATE: Illinois
;      COUNTRY: USA
;      ZIP: 60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/480,551
;      FILING DATE:
;      CLASSIFICATION: 530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: US 08/264,176
;      FILING DATE:
;      APPLICATION NUMBER: US 07/653,012
;      FILING DATE: 08-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME: McDonnell, John J
;      REGISTRATION NUMBER: 26,949
;      REFERENCE/DOCKET NUMBER: 90,1104
;    TELECOMMUNICATION INFORMATION:

;      TELEPHONE: 312 715 1000
;      TELEFAX: 312 715 1234
;      TELEX: 910-221-5317
;  INFORMATION FOR SEQ ID NO: 13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 18 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
US-08-480-551-13

Query Match             100.0%;  Score 31;  DB 2;  Length 18;
Best Local Similarity   100.0%;  Pred. No. 4.4;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PGVGVA 6
              ||||||
Db          5 PGVGVA 10

RESULT 14
US-08-612-842-14
; Sequence 14, Application US/08612842
; Patent No. 5976495
; GENERAL INFORMATION:
;    APPLICANT: POLLAK, ALFRED
;    APPLICANT: GOODBODY, ANNE
;    TITLE OF INVENTION: PEPTIDE DERIVED RADIONUCLIDE CHELATORS
;    NUMBER OF SEQUENCES: 39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: NIKAIDO, MARMELSTEIN, MURRAY & ORAM
;      STREET: 655 15TH STREET, NW, G STREET LOBBY, SUITE
;      STREET: 330
;      CITY: WASHINGTON
;      STATE: DC
;      COUNTRY: USA
;      ZIP: 20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/612,842
;      FILING DATE: 20-MAR-1996
;      CLASSIFICATION: 424
;    ATTORNEY/AGENT INFORMATION:
;      NAME: BERMAN, RICHARD J
;      REGISTRATION NUMBER: 39107
;      REFERENCE/DOCKET NUMBER: 8012-6002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 202 638-5000
;      TELEFAX: 202 638-4810
;  INFORMATION FOR SEQ ID NO: 14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 18 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
US-08-612-842-14

Query Match             100.0%;  Score 31;  DB 2;  Length 18;
Best Local Similarity   100.0%;  Pred. No. 4.4;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PGVGVA 6
              ||||||
Db          5 PGVGVA 10
```

RESULT 15
US-08-290-853-28
; Sequence 28, Application US/08290853
; Patent No. 5989519
; GENERAL INFORMATION:
; APPLICANT: Dean, Richard T
; APPLICANT: Buttram, Scott
; TITLE OF INVENTION: Technetium-99m Labeled Peptides for
; TITLE OF INVENTION: Imaging Inflammation
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,853
; FILING DATE: 11-OCT-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5989519nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 92,112-H
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1..3
; OTHER INFORMATION: /label= BAT
; OTHER INFORMATION: /note= "The amino terminal residue is
; OTHER INFORMATION: linked to a BAT radiolabel binding moiety."
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 18
; OTHER INFORMATION: /label= Amide
; OTHER INFORMATION: /note= "The carboxyl terminus is modified to an
; OTHER INFORMATION: amide"
US-08-290-853-28

Query Match   100.0%; Score 31; DB 2; Length 18;
Best Local Similarity   100.0%; Pred. No. 4.4;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
          ||||||
Db   5 PGVGVA 10

Search completed: April 24, 2001, 16:36:24
Job time: 304 sec

Copyright GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:41:58 ; Search time 74.56 Seconds
                 (without alignments).
                 5.530 Million cell updates/sec

Title: US-09-340-736-5
Perfect score: 31
Sequence: 1 PGVGVA 6

Scoring table: BLOSUM62
                 GapOp 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR_67:*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	31	100.0	76	2	I45885	elastin - bovine (
2	31	100.0	747	1	EABO	elastin precursor,
3	31	100.0	770	2	S59623	tropoelastin - she
4	31	100.0	792	1	EAHU	elastin precursor,
5	31	100.0	907	2	A45560	sporozoite surface
6	31	100.0	1146	2	B35962	protein-tyrosine k
7	31	100.0	1174	2	A42459	parasporal crystal
8	31	100.0	1174	2	S32649	parasporal crystal
9	31	100.0	1182	2	A35962	protein-tyrosine k
10	30	96.8	358	2	T45934	hypothetical prote
11	30	96.8	456	1	T34955	probable cell diri
12	30	96.8	634	2	C83530	potassium uptake p
13	30	93.5	646	2	A70903	probable exoinucle
14	30	93.5	216	2	S73913	polypeptide deform
15	29	93.5	226	1	G64211	formylmethionine d
16	29	93.5	317	2	T35981	probable peptide t
17	29	93.5	390	2	A72108	hypothetical prote
18	29	93.5	617	2	T49059	hypothetical prote
19	29	90.3	85	1	S18314	cytochrome-c oxida
20	28	90.3	85	1	S01156	cytochrome-c oxida
21	28	90.3	85	1	T43006	cytochrome-c oxida
22	28	90.3	105	2	F77030	hypothetical prote
23	28	90.3	112	2	S52088	cytochrome-c oxida
24	28	90.3	182	1	B83982	formylmethionine d
25	28	90.3	184	1	D69862	formylmethionine d
26	28	90.3	312	2	T27004	hypothetical prote
27	28	90.3	368	1	I39598	N-Acylamino acid r
28	28	90.3	501	2	T04735	cytochrome P450 ho
29	28	90.3	543	2	T00513	cytochrome P450 ho
30	28	90.3	588	2	A53563	synaptotagmin III
31	28	90.3	609	2	T32456	hypothetical prote
32	28	90.3	623	2	S54072	probable membrane
33	28	90.3	651	2	A26581	beta-glucuronidase
34	28	90.3	690	2	S28222	peroxidase (EC 1.1
35	28	90.3	777	1	TVVPCP	large T antigen -
36	28	90.3	782	2	S22560	large T antigen -
37	28	90.3	785	1	TVVPT	large T antigen -
38	28	90.3	856	1	T10585	serine proteinase
39	28	90.3	1017	2	A37227	Na+/K+-exchanging
40	28	90.3	1020	2	B24639	Na+/K+-exchanging
41	28	90.3	1020	1	A34474	Na+/K+-exchanging
42	28	90.3	1021	1	S04630	Na+/K+-exchanging
43	28	90.3	1021	1	PWSHNA	Na+/K+-exchanging
44	28	90.3	1021	2	B24862	Na+/K+-exchanging
45	28	90.3	1023	2	A24414	Na+/K+-exchanging

ALIGNMENTS

RESULT   1
I45885
elastin - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change 13-Aug-1999
C;Accession: I45885
R;Rosenbloom, J.
Lab. Invest. 51, 605-623, 1984
A;Title: Biology of disease: Elastin: Relation of protein and gene structure to disea
A;Reference number: I45885; MUID:85059254
A;Accession: I45885
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Cross-references: GB:M18991; NID:g163008; PIDN:AAA96416.1; PID:g552319
C;Genetics:
A;Introns: 20/1; 58/1
C;Superfamily: elastin

Query Match   100.0%; Score 31; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 14;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
            ||||||
Db   23 PGVGVA 28

RESULT   2
EABO
elastin precursor, splice form a - bovine
N;Alternate names: tropoelastin
N;Contains: elastin precursor, splice form b; elastin precursor, splice form c
C;Species: Bos primigenius taurus (cattle)
C;Date: 08-Jun-1989 #sequence_revision 26-Jul-1996 #text_change 22-Jun-1999
C;Accession: A31865; A26720; B26728; C26728; A22343; I45886
R;Yeh, H.; Anderson, N.; Ornstein-Goldstein, N.; Bashir, M.M.; Rosenbloom, J.C.; Abra
Biochemistry 28, 2365-2370, 1989
A;Title: Structure of the bovine elastin gene and S1 nuclease analysis of alternative
A;Reference number: A31865; MUID:89274159
A;Accession: A31865
A;Molecule type: DNA
A;Residues: 1-27 <EH>
A;Cross-references: GB:J02855; NID:g340504; PIDN:AAA30776.1; PID:g552339
R;Raju, K.; Anwar, R.A.
J. Biol. Chem. 262, 5755-5762, 1987
A;Title: Primary structures of bovine elastin a, b, and c deduced from the sequences
A;Reference number: A26728
A;Accession: A26728
A;Molecule type: mRNA
A;Residues: 1,'RS',4-11,'E',13-636,'V',638-747 <RAJ>

A.;Cross-references: GB:J02717; NID:g163019; PIDN:AAA30503.1; PID:g163020
A.;Accession: B26728
A.;Molecule type: mRNA
A.;Residues: 1,'RS',4-11,'E',13-225,240-636,'V',638-747 <RA2>
A.;Cross-references: GB:K03505; NID:g163025; PIDN:AAA30505.1; PID:g163026
A.;Accession: C26728
A.;Molecule type: mRNA
A.;Residues: 1,'RS',4-11,'E',13-225,260-636,'V',638-747 <RA3>
A.;Cross-references: GB:K03506; NID:g163027; PIDN:AAA30506.1; PID:g163028
R;Cicila, G.; May, M.; Ornstein-Goldstein, N.; Indik, Z.; Yeh, H.S.; Rosenbl
Biochemistry 24, 3075-3080, 1985
A.;Title: Structure of the 3' portion of the bovine elastin gene.
A.;Reference number: A22343; MUID:85280426
A.;Accession: A22343
A.;Molecule type: DNA
A.;Residues: 613-747 <CIC>
A.;Cross-references: GB:M20415
R;Rosenbloom, J.
Lab. Invest. 51, 605-623, 1984
A.;Title: Biology of disease: Elastin: Relation of protein and gene structure to disease.
A.;Reference number: I45885; MUID:85059254
A.;Accession: I45886
A.;Molecule type: DNA
A.;Residues: 678-683,685-747 <ROS>
A.;Cross-references: GB:M11898; NID:g163015, PIDN:AAA96417.1; PID:g163018
R;Brown, P.L.; Mecham, L.; Tisdale, C.; Mecham, R.P.
Biochem. Biophys. Res. Commun 186, 549-555, 1992
A.;Title: The cysteine residues in the carboxy terminal domain of tropoelastin form an in
A.;Reference number: A58621; MUID:92337651
A.;Contents: annotation. disulfide bonds
C;Comment: The term tropoelastin refers to a soluble precursor form of the extracellular
ine oxidase activity.
C;Genetics:
A.;Introns: 634/3; 653/3; 676/3; 689/3; 707/3; 716/3; 733/3
A.;Note: the list of introns is incomplete
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;1-747/Product: elastin precursor, splice form a #status predicted <EPA>
F;1-225,260-747/Product: elastin precursor, splice form c #status predicted <EPC>
F;1-225,240-747/Product: elastin precursor, splice form b #status predicted <EPB>
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-747/Product: elastin #status predicted <MAT>
F;105,109,252,271,275,324,327,400,404,407,445,448,489,493,544,548,552,606,609,645,649,68
F;737-742/Disulfide bonds: #status experimental

Query Match   100.0%; Score 31; DB 1; Length 747;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches   6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

      1 PGVGVA 6
        ||||||
Db 501 PGVGVA 506

RESULT 3
S59623
tropoelastin - sheep
C;Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 23-Aug-1996 #sequence_revision 13-Mar-1997 #text_change 24-Oct-1997
C;Accession: S59623; A24758
R;Mauch, J.C.; Sandberg, L.B.; Roos, P.J.; Jimenez, F.; Christiano, A.M.; Deak, S.B.; Bo
Matrix Biol 14, 535-641, 1994
A.;Title: Extensive alternate exon usage at the 5' end of the sheep tropoelastin gene.
A.;Reference number: S59623
A.;Status: preliminary; not compared with conceptual translation
A.;Molecule type: mRNA
A.;Residues: 1-770 <MAU>
R;Yoon, K.; Davidson, J.M.; Boyd, C.; May, M.; LuValle, P.; Ornstein-Goldstein, N.; Smit
Arch. Biochem. Biophys. 241, 684-691, 1985
A.;Title: Analysis of the 3' region of the sheep elastin gene.

A.;Reference number: A24758; MUID:85305763
A.;Accession: A24758
A.;Molecule type: mRNA
A.;Residues: 655-669,671-716,732-770 <YOO>
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;760-765/Disulfide bonds: #status predicted

Query Match   100.0%; Score 31; DB 2; Length 770;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches   6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 PGVGVA 6
        ||||||
Db 500 PGVGVA 505

RESULT 4
EAHU
elastin precursor, long splice form - human
N;Alternate names: tropoelastin
C;Species: Homo sapiens (man)
C;Date: 22-Jun-1990 #sequence_revision 26-Jul-1996 #text_change 22-Jun-1999
C;Accession: A32707; A33705; A30524; A53891
R;Indik, Z.; Yeh, H.; Ornstein-Goldstein, N.; Sheppard, P.; Anderson, N.; Rosenbloom,
Proc Natl. Acad. Sci. U.S.A. 84, 5580-5584, 1987
A.;Title: Alternative splicing of human elastin mRNA indicated by sequence analysis of
A.;Reference number: A32707; MUID:87289668
A.;Accession: A32707
A.;Molecule type: mRNA
A.;Residues: 1-500,507-792 <IND>
R;Bashir, M.M.; Indik, Z.; Yeh, H.; Ornstein-Goldstein, N.; Rosenbloom, J.C.; Abrams,
J. Biol. Chem. 264, 8887-8891, 1989
A.;Title: Characterization of the complete human elastin gene. Delineation of unusual
A.;Reference number: A33705; MUID:89255358
A.;Accession: A33705
A.;Molecule type: DNA
A.;Residues: 1-27 <BAS>
A.;Cross-references: GB:J04821; NID:g182052; PIDN:AAA52379.1; PID:g553276
R;Fazio, M.J.; Olsen, D.R.; Rauh, E.A.; Baldwin, C.T.; Indik, Z.; Ornstein-Goldstein,
J. Invest. Dermatol 91, 458-464, 1988
A.;Title: Cloning of full-length elastin cDNAs from a human skin fibroblast recombinan
A.;Reference number: A30524; MUID:89009960
A.;Accession: A30524
A.;Molecule type: mRNA
A.;Residues: 1-453,483-617,651-792 <FAZ>
A.;Cross-references: EMBL:M36860; NID:g182061; PIDN:AAA52382.1; PID:g182062
A.;Note: this sequence represents a composite of several splice forms
R;Fazio, M.J.; Olsen, D.R.; Kuivaniemi, H.; Chu, M.L.; Davidson, J.M.; Rosenbloom, J.
Lab. Invest. 58, 270-277, 1988
A.;Title: Isolation and characterization of human elastin cDNAs, and age-associated va
A.;Reference number: A53891; MUID:88156138
A.;Accession: A53891
A.;Molecule type: mRNA
A.;Residues: 164-453,483-500,507-617,651-792 <FA2>
A.;Cross-references: GB:M24782; NID:g182063; PIDN:AAA53190.1; PID:g182064
C;Comment: The term tropoelastin refers to a soluble precursor form of the extracellu
ine oxidase activity.
C;Genetics:
A.;Gene: GDB:ELN
A.;Cross-references: GDB:119107; OMIM:130160
A.;Map position: 7q11.23-7q11.23
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-792/Product: elastin #status predicted <MAT>
F;782-787/Disulfide bonds: #status predicted

Query Match   100.0%; Score 31; DB 1; Length 792;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1 PGVGVA 6
         ||||||
Db   510 PGVGVA 515
```

RESULT 5
A45560
sporozoite surface antigen SPAG-1 - Theileria annulata
C;Species: Theileria annulata
C;Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C;Accession: A45560
R;Hall, R.; Hunt, P.D.; Carrington, M.; Simmons, D.; Williamson, S.; Mecham, R.P.; Tait,
Mol. Biochem. Parasitol. 53, 105-112, 1992
A;Title: Mimicry of elastin repetitive motifs by Theileria annulata sporozoite surface a
A;Reference number: A45560; MUID:92365719
A;Accession: A45560
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-907 <HAL>
A;Cross-references: GB:M63017; NID:g161884; PID:g161885
A;Note: sequence extracted from NCBI backbone (NCBIN:111148, NCBIP:111150)
C;Keywords: surface antigen

Query Match 100.0%; Score 31; DB 2; Length 907;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1 PGVGVA 6
         ||||||
Db   215 PGVGVA 220
```

RESULT 6
B35962
protein-tyrosine kinase (EC 2.7.1.112) I, splice form A - human
N;Alternate names: c-abl proto-oncogene homolog 2
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 04-Feb-2000
C;Accession: B35962
R;Kruh, G.D.; Perego, R.; Miki, T.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 87, 5802-5806, 1990
A;Title: The complete coding sequence of arg defines the Abelson subfamily of cytoplasmi
A;Reference number: A35962; MUID:90332670
A;Accession: B35962
A;Molecule type: mRNA
A;Residues: 1-1146 <KRU>
A;Cross-references: GB:M35296
C;Genetics:
A;Gene: GDB:ABL2; ABLL
A;Cross-references: GDB:119641; OMIM:164690
A;Map position: 1q24-1q25
C;Superfamily: human protein-tyrosine kinase abl; protein kinase homology; SH2 homology;
C;Keywords: alternative splicing; ATP; kinase-related transforming protein; phosphotranse
F;78-126/Domain: SH3 homology <SH3>
F;137-227/Domain: SH2 homology <SH2>
F;250-510/Domain: protein kinase homology <KIN>
F;258-265/Region: protein kinase ATP-binding motif

Query Match 100.0%; Score 31; DB 2; Length 1146;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1 PGVGVA 6
         ||||||
Db   841 PGVGVA 846
```

RESULT 7
A42459

---

parasporal crystal protein cryIFa1 - Bacillus thuringiensis (strain aizawai)
N;Alternate names: parasporal crystal protein cryIF
C;Species: Bacillus thuringiensis
C;Date: 10-Jul-1992 #sequence_revision 10-Jul-1992 #text_change 01-Dec-2000
C;Accession: A42459
R;Chambers, J.A.; Jelen, A.; Gilbert, M.P.; Jany, C.S.; Johnson, T.B.; Gawron-Burke,
J. Bacteriol. 173, 3966-3976, 1991
A;Title: Isolation and characterization of a novel insecticidal crystal protein gene
A;Reference number: A42459; MUID:91286178
A;Accession: A42459
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1174 <CHA>
A;Cross-references: GB:M63897; NID:g142757; PIDN:AAA22348.1; PID:g142758
C;Superfamily: parasporal crystal protein
C;Keywords: delta-endotoxin

Query Match 100.0%; Score 31; DB 2; Length 1174;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1 PGVGVA 6
         ||||||
Db    50 PGVGVA 55
```

RESULT 8
S32649
parasporal crystal protein cryIFa3 - Bacillus thuringiensis
C;Species: Bacillus thuringiensis
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 01-Dec-2000
C;Accession: S32649
R;Lambert, B.
submitted to the EMBL Data Library, April 1993
A;Reference number: S32645
A;Accession: S32649
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1174 <LAM>
A;Cross-references: EMBL:Z22512; NID:g295865; PIDN:CAA80235.1; PID:g295866
C;Superfamily: parasporal crystal protein
C;Keywords: delta-endotoxin

Query Match 100.0%; Score 31; DB 2; Length 1174;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1 PGVGVA 6
         ||||||
Db    50 PGVGVA 55
```

RESULT 9
A35962
protein-tyrosine kinase (EC 2.7.1.112) I, splice form B - human
N;Alternate names: c-abl proto-oncogene homolog 2
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 04-Feb-2000
C;Accession: A35962; A47577
R;Kruh, G.D.; Perego, R.; Miki, T.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 87, 5802-5806, 1990
A;Title: The complete coding sequence of arg defines the Abelson subfamily of cytopla
A;Reference number: A35962; MUID:90332670
A;Accession: A35962
A;Molecule type: mRNA
A;Residues: 1-1182 <KRU>
A;Cross-references: GB:M35296; NID:g178992; PIDN:AAA35553.1; PID:g178993
R;Kruh, G.D.; King, C.R.; Kraus, M.H.; Popescu, N.C.; Amsbaugh, S.C.; McBride, W.O.;
Science 234, 1545-1548, 1986
A;Title: A novel human gene closely related to the abl proto-oncogene.
A;Reference number: A47577; MUID:87069952

A;Accession: A47577
A;Status: preliminary: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 'TI',345-469 <KRU2>
A;Cross-references: GB;M14903; NID:g177954; GB:M14904; NID:g177955
C;Genetics:
A;Gene: GDB;ABL2; ABL1
A;Cross-references: GDB:119641; OMIM:164690
C;Superfamily: human protein-tyrosine kinase abl; protein kinase homology: SH2 homology:
C;Keywords: alternative splicing; ATP; kinase-related transforming protein: phosphotrans
F;114-162/Domain: SH3 homology <SH3>
F;173-263/Domain: SH2 homology <SH2>
F;286-546/Domain: protein kinase homology <KIN>
F;294-302/Region: protein kinase ATP-binding motif

Query Match 100.0%; Score 31; DB 2; Length 1182;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
  |||||
Db 877 PGVGVA 882

RESULT 10
T45934
hypothetical protein F5K20.240 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 17-Mar-2000
A;Accession: T45934
R;Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23017
A;Accession: T45934
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-358 <MON>
A;Cross-references: EMBL:AL132960
A;Experimental source: cultivar Columbia: BAC clone F5K20
C;Genetics:
A;Map position: 3
A;Introns: 98/3; 157/3; 207/3; 234/3; 258/3; 281/1
C;Note: F5K20.240
C;Superfamily: ADP,ATP carrier protein: ADP,ATP carrier protein repeat homology;

Query Match 96.8%; Score 30; DB 2; Length 358;
Best Local Similarity 83.3%; Pred. No. 1e-02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
  |||:|
Db 337 PGVGVR 342

RESULT 11
T34955
probable cell division protein ftsW - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 03-Dec-1999
A;Accession: T34955
R;Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A;Reference number: 221563
A;Accession: T34955
A;Status: preliminary: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-456 <SAU>
A;Cross-references: EMBL:AL109663; PIDN:CAB51994.1; GSPDB:GN00070; SCOEDB:SC4A10.18c
A;Experimental source: strain A3(2)
C;Genetics:

A;Gene: ftsW; SCOEDB:SC4A10.18c

Query Match 96.8%; Score 30; DB 2; Length 456;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
  ||:|||
Db 133 PGIGVA 138

RESULT 12
CB3530
potassium uptake protein Kup PA0917 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
A;Accession: CB3530
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa
Nature 406, 959-964, 2000
A;Reference number: AB2950; MDID:20437337
A;Accession: CB3530
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-634 <STO>
A;Cross-references: GB:AE004526; GB:AE004091; NID:g9946819; PIDN:AAG04306.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: kup; PA0917

Query Match 96.8%; Score 30; DB 2; Length 634;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
  ||:|||
Db 223 PGIGVA 228

RESULT 13
A70903
probable excinuclease abc subunit c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
A;Accession: A70903
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squres, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MDID:98295987
A;Accession: A70903
A;Status: preliminary: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-646 <COL>
A;Cross-references: GB:Z80108; GB:AL123456; NID:g3256012; PIDN:CAB02168.1; PID:g15429
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: uvrC
C;Superfamily: excinuclease ABC chain C

Query Match 96.8%; Score 30; DB 2; Length 646;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PGVGVA 6
  ||:||

Db   623 PGIGVA 628

RESULT 14
S73913
polypeptide deformylase def - Mycoplasma pneumoniae (strain ATCC 29342)
N;Alternate names: hypothetical protein K04_orf216
C;Species: Mycoplasma pneumoniae
A;Variety: ATCC 29342
C;Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C;Accession: S73913
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A;Reference number: S73327; MUID:97105885
A;Accession: S73913
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-216 <HIM>
A;Cross-references: EMBL:AE000057; GB:U00089; NID:g1674279; PIDN:AAB96235.1; PID:g1674278
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C;Genetics:
A;Gene: def
A;Genetic code: SGC3
C;Superfamily: polypeptide deformylase

Query Match   93.5%; Score 29; DB 2; Length 215;
Best Local Similarity   66.7%; Pred. No. 1e+02;
Matches   4; Conservative   2; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
            ||:||:|
Db   82 PGIGIA 87

RESULT 15
G64211
formylmethionine deformylase homolog - Mycoplasma genitalium
C;Species: Mycoplasma genitalium
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999
C;Accession: G64211
R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J
, C.A.; Venter, J.C.
Science 270, 397-403, 1995
A;Title: The minimal gene complement of Mycoplasma genitalium.
A;Reference number: A64200; MUID:96026346
A;Accession: G64211
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-226 <TIGR>
A;Cross-references: GB:U39590; GB:L43967; NID:g1045782; PID:g1045785; TIGR:MG106
A;Experimental source: strain G-37
C;Genetics:
A;Genetic code: SGC3
A;Start codon: GTG
C;Superfamily: polypeptide deformylase

Query Match   93.5%; Score 29; DB 1; Length 226;
Best Local Similarity   66.7%; Pred. No. 1e+02;
Matches   4; Conservative   2; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
            ||:||:|
Db   92 PGIGIA 97

Search completed: April 24, 2001, 16:42:01
Job time: 465 sec

```
                    GenCore version 4.5
              Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search. using sw model

Run on:        April 24, 2001, 16:42:48 ; Search time 44.88 Seconds
                                          (without alignments)
                                          4.580 Million cell updates/sec

Title:         US-09-340-736-5
Perfect score: 31
Sequence:      1 PGVGVA 6

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Matched:       93435 seqs, 31255436 residues

Total number of hits satisfying chosen parameters:        93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 listing first 45 summaries

Database :     SwissProt_39:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                              SUMMARIES

Result          %     Query
No.    Score  Match  Length DB   ID            Description
------ -----  -----  ------ --   ----------    --------------------
   1    31   100.0    730   1    ELS_HUMAN     P15502 homo sapien
   2    31   100.0    747   1    ELS_BOVIN     P04985 bos taurus
   3    31   100.0   1169   1    C1FB_BACTM    O56377 bacillus th
   4    31   100.0   1174   1    C1FA_BACTA    Q03746 bacillus th
   5    31   100.0   1182   1    ABL2_HUMAN    P42684 homo sapien
   6    30    96.8    546   1    UVRC_MYCTU    P71689 mycobacteri
   7    29    93.5    216   1    DEF_MYCGE     P47352 mycoplasma
   8    29    93.5    216   1    DEF_MYCPN     P75527 mycoplasma
   9    28    90.3     85   1    COXE_BOVIN    P13182 bos taurus
  10    28    90.3    105   1    YK74_MYCTU    Q50689 mycobacteri
  11    28    90.3    111   1    COXE_MOUSE    P43024 mus musculu
  12    28    90.3    111   1    COXE_RAT      P10818 rattus norv
  13    28    90.3    184   1    DEF2_BACSU    Q45495 bacillus su
  14    28    90.3    292   1    YI52_HUMAN    Q14165 homo sapien
  15    28    90.3    430   1    IDHP_CANTR    O13285 candida tro
  16    28    90.3    431   1    GUDH_STRCO    Q9Zde9 streptomyce
  17    28    90.3    587   1    SYT3_MOUSE    O35681 mus musculu
  18    28    90.3    588   1    SYT3_RAT      P40748 rattus norv
  19    28    90.3    648   1    BGLR_CERAE    O77695 cercopithec
  20    28    90.3    651   1    BGLR_HUMAN    P08236 homo sapien
  21    28    90.3    690   1    PERO_DROME    Q01603 drosophila
  22    28    90.3    777   1    TALA_POVMC    P12905 mouse polyo
  23    28    90.3    782   1    TALA_POVM3    P03074 mouse polyo
  24    28    90.3    785   1    TALA_POVMA    P03073 mouse polyo
  25    28    90.3   1017   1    ACN2_CHICK    P24797 gallus gall
  26    28    90.3   1020   1    ATN2_HUMAN    P50993 homo sapien
  27    28    90.3   1020   1    ATN2_RAT      P06686 rattus norv
  28    28    90.3   1021   1    ATN1_CANFA    P50097 canis famil
  29    28    90.3   1021   1    ATN1_HORSE    P18907 equus cabal
  30    28    90.3   1021   1    ATN1_PIG      P05024 sus scrofa
  31    28    90.3   1021   1    ATN1_SHEEP    P04074 ovis aries
  32    28    90.3   1021   1    ATN3_PIG      P18874 sus scrofa
  33    28    90.3   1023   1    ATN1_HUMAN    P05023 homo sapien
```

```
  34    27    87.1     34     ATPL_STRCO    P50014 streptomyce
  35    27    87.1     35     COXE_CYPCA    O13082 cyprinus ca
  36    27    87.1     36     VGLL_PRVKA    P52512 pseudorabie
  37    27    87.1     37     DEF2_BACST    O31410 bacillus st
  38    27    87.1     38     HIS5_SALTY    P10376 salmonella
  39    27    87.1     39     HIS2_CAMJE    Q9pm75 campylobact
  40    27    87.1     40     HIS5_BUCAI    P57204 buchnera ap
  41    27    87.1     41     HIS5_ECOLI    P10375 escherichia
  42    27    87.1     42     HIS5_BUCAP    Q92be3 buchnera ap
  43    27    87.1     43     HIS5_HAEIN    P44310 haemophilus
  44    27    87.1     44     HIS5_VIBCH    Q9ksx0 vibrio chol
  45    27    87.1     45     FLAH_METVO    O06641 methanococc
```

```
                                        ALIGNMENTS

RESULT 1
ELS_HUMAN
ID   ELS_HUMAN        STANDARD;       PRT;    730 AA.
AC   P15502;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN).
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin fibroblast;
RX   MEDLINE=89009960; PubMed=3171121;
RA   Fazio M.J., Olsen D.R., Kauh E.A., Baldwin C.T., Indik Z.,
RA   Ornstein-Goldstein N., Yeh H., Rosenbloom J., Uitto J.;
RT   "Cloning of full-length elastin cDNAs from a human skin fibroblast
RT   recombinant cDNA library: further elucidation of alternative splicing
RT   utilizing exon-specific oligonucleotides.";
RL   J. Invest. Dermatol. 91:458-464(1988).
RN   [2]
RP   SEQUENCE OF 603-730 FROM N.A.
RC   TISSUE=Placenta, and Hippocampus;
RX   MEDLINE=96291399; PubMed=8689688;
RA   Franciskakis J.M., Ewart A.K., Morris C.A., Mervis C.B.,
RA   Bertrand J., Robinson B.F., Klein B.P., Ensing G.J., Everett L.A.,
RA   Green E.D., Proeschel C., Gutowski N.J., Noble M., Atkinson D.L.,
RA   Odelberg S.J., Keating M.T.;
RT   "LIM-kinase1 hemizygosity implicated in impaired visuospatial
RT   constructive cognition.";
RL   Cell 86:159-69(1996).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   -!- DISEASE: DELETED IN WILLIAMS-BEUREN SYNDROME (WBS), A
CC       DEVELOPMENTAL DISORDER CAUSED BY HAPLOINSUFFICIENCY FOR GENES IN A
CC       2-CM REGION OF CHROMOSOME BAND 7Q11.23. HEMIZYGOUS DELETION MAY
CC       CONTRIBUTE TO CERTAIN DEFECTS SUCH AS HYPERCALCEMIA AND GROWTH
CC       DELAY
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics  and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on  its
CC   use  by  non-profit  institutions as long  as its content is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X15603; CAA33627.1; -
DR   EMBL; M36860; AAA52382.1; -
```

```
DR   EMBL; U62292; AAB17544.1; -.
DR   PIR; A30524; A30524.
DR   HSSP; P04002; 1WFA.
DR   MIM; 130160; -.
DR   MIM; 194050; -.
KW   Structural protein; Connective tissue; Repeat; Signal.
FT   SIGNAL        1     26
FT   CHAIN        27    730       ELASTIN.
SQ   SEQUENCE    730 AA;  63260 MW;  AB06D15BA567AE46 CRC64;

Query Match             100.0%;  Score 31;  DB 1;  Length 730;
Best Local Similarity   100.0%;  Pred. No. 72;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PGVGVA 6
              ||||||
Db        481 PGVGVA 486

RESULT 2
ELS_BOVIN     STANDARD;      PRT;    747 AA.
P04985; P04986; P04987; Q29421;
13-AUG-1987 (Rel. 05, Created)
13-AUG-1987 (Rel. 05, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
ELASTIN PRECURSOR (TROPOELASTIN).
ELN.
Bos taurus (Bovine).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bovinae; Bos.
NCBI_TaxID=9913;
[1]
SEQUENCE FROM N.A.
TISSUE=Nuchal ligament;
MEDLINE=87194772; PubMed=3032943;
Raju K., Anwar R.A.;
"Primary structures of bovine elastin a, b, and c deduced from the
sequences of cDNA clones.";
J. Biol. Chem. 262:5755-5762(1987).
[2]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   TISSUE=Nuchal ligament;
RX   MEDLINE=89274159; PubMed=2543440;
RA   Yeh H., Anderson N., Ornstein-Goldstein N., Bashir M.M.,
RA   Rosenbloom J.C., Abrams W.R., Indik Z., Yoon K., Parks W., Mecham R.,
RA   Rosenbloom J.;
RT   "Structure of the bovine elastin gene and S1 nuclease analysis of
RT   alternative splicing of elastin mRNA in the bovine nuchal ligament.";
RL   Biochemistry 28:2365-2370(1989).
[3]
RP   SEQUENCE OF 1-27 FROM N.A.
RX   MEDLINE=91234332; PubMed=2031719;
RA   Manohar A., Shi W., Anwar R.A.;
RT   "Partial characterization of bovine elastin gene; comparison with the
RT   gene for human elastin.";
RL   Biochem. Cell Biol. 59:185-192(1991).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS: A (SHOWN HERE), B AND C; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; J02717; AAA30503.1; -.
DR   EMBL; K03505; AAA30505.1; -.
DR   EMBL; K03506; AAA30506.1; -.
DR   EMBL; J02855; AAA30776.1; -.
DR   EMBL; M58652; AAA03519.1; -.
DR   PIR; A26728; A26728.
DR   PIR; B26728; B26728.
DR   PIR; C26728; C26728.
DR   HSSP; P04002; 1WFA.
KW   Structural protein; Connective tissue; Repeat; Signal;
KW   Alternative splicing
FT   SIGNAL        1     26
FT   CHAIN        27    747       ELASTIN.
FT   MOD_RES     105    105       OXIDATIVE DEAMINATION.
FT   MOD_RES     109    109       OXIDATIVE DEAMINATION.
FT   MOD_RES     252    252       OXIDATIVE DEAMINATION.
FT   MOD_RES     271    271       OXIDATIVE DEAMINATION.
FT   MOD_RES     275    275       OXIDATIVE DEAMINATION.
FT   MOD_RES     324    324       OXIDATIVE DEAMINATION.
FT   MOD_RES     327    327       OXIDATIVE DEAMINATION.
FT   MOD_RES     400    400       OXIDATIVE DEAMINATION.
FT   MOD_RES     404    404       OXIDATIVE DEAMINATION.
FT   MOD_RES     407    407       OXIDATIVE DEAMINATION.
FT   MOD_RES     448    448       OXIDATIVE DEAMINATION.
FT   MOD_RES     489    489       OXIDATIVE DEAMINATION.
FT   MOD_RES     493    493       OXIDATIVE DEAMINATION.
FT   MOD_RES     544    544       OXIDATIVE DEAMINATION.
FT   MOD_RES     548    548       OXIDATIVE DEAMINATION.
FT   MOD_RES     552    552       OXIDATIVE DEAMINATION.
FT   MOD_RES     606    606       OXIDATIVE DEAMINATION.
FT   MOD_RES     609    609       OXIDATIVE DEAMINATION.
FT   MOD_RES     645    645       OXIDATIVE DEAMINATION.
FT   MOD_RES     649    649       OXIDATIVE DEAMINATION.
FT   MOD_RES     685    685       OXIDATIVE DEAMINATION.
FT   MOD_RES     688    688       OXIDATIVE DEAMINATION.
FT   VARSPLIC    226    239       MISSING (IN ELASTIN B).
FT   VARSPLIC    226    259       MISSING (IN ELASTIN C).
FT   CONFLICT     2      3        RS -> AG (IN REF. 2 AND 3).
FT   CONFLICT    12     12        E -> G (IN REF. 2 AND 3).
SQ   SEQUENCE    747 AA;  64229 MW;  633C03E411643D83 CRC64;

Query Match             100.0%;  Score 31;  DB 1;  Length 747;
Best Local Similarity   100.0%;  Pred. No. 74;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PGVGVA 6
              ||||||
Db        501 PGVGVA 506

RESULT 3
CIFB_BACTM    STANDARD;      PRT;   1169 AA.
Q9RC19;
066377; Q9RC19;
30-MAY-2000 (Rel. 39, Created)
30-MAY-2000 (Rel. 39, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
PESTICIDAL CRYSTAL PROTEIN CRY1FB (INSECTICIDAL DELTA-ENDOTOXIN
CRY1F(B)) (CRYSTALINE ENTOMOCIDAL PROTOXIN) (132 KDA CRYSTAL PROTEIN).
CRY1FB OR CRY1F(B) OR CRYINA67-1.
Bacillus thuringiensis (subsp. morrisoni).
Bacillus thuringiensis (subsp. morrisoni).
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
NCBI_TaxID=1441;
[1]
SEQUENCE FROM N.A.
Song F., Zhang J., Ding Z., Chen Z., Li G., Huang D.;
"A novel cry1Fb gene from Bacillus thuringiensis subsp. morrisoni.";
Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
[2]
```

RP   SEQUENCE FROM N.A.
RC   STRAIN=INA67;
RA   Masuda K., Asano S.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROMOTES COLLOIDOSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC   EPITHELIAL CELLS OF INSECTS.
CC   -!- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC   SPORULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC   OF THE SPORE COAT.
CC   -!- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC   N-TERMINUS.
CC   -!- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC   -!- --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -!- --
DR   EMBL; AF063350; AAF21767.1; -.
DR   EMBL; AB012288; BAA25298.1; -.
DR   HSSP; P02965; 1CIY.
DR   InterPro; IPR001178; -.
DR   Pfam; PF00555; endotoxin; 1.
KW   Toxin; Sporulation.
FT   CONFLICT   338   338   V -> I (IN REF. 2).
FT   CONFLICT   382   382   G -> A (IN REF. 2).
FT   CONFLICT   1002   1002   V -> I (IN REF. 2).
FT   CONFLICT   1012   1012   E -> K (IN REF. 2).
FT   CONFLICT   1053   1053   MISSING (IN REF. 2).
FT   CONFLICT   1085   1085   G -> R (IN REF. 2).
SQ   SEQUENCE   1169 AA; 132284 MW; EFCAAF6F6EE73C1B CRC64;

Query Match   100.0%; Score 31; DB 1; Length 1169;
Best Local Similarity   100.0%; Pred. No. 1.1e+02;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
              ||||||
Db   50 PGVGVA 55

ULT   4
_BACTA
ID   CI1FA_BACTA   STANDARD;   PRT;   1174 AA.
AC   Q03746;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PESTICIDAL CRYSTAL PROTEIN CRY1FA (INSECTICIDAL DELTA-ENDOTOXIN
DE   CRY1F(A)) (CRYSTALINE ENTOMOCIDAL PROTOXIN) (134 KDA CRYSTAL PROTEIN).
GN   CRY1FA OR CRY1F(A) OR CRY1F.
OS   Bacillus thuringiensis (subsp. aizawai).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1433;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EG6346;
RX   MEDLINE=91286178; PubMed=2061280;
RA   Chambers J.A., Jelen A., Gilbert M.P., Jany C.S., Johnson T.B.,
RA   Gawron-Burke C.;
RT   "Isolation and characterization of a novel insecticidal crystal
RT   protein gene from Bacillus thuringiensis subsp. aizawai.";
RL   J. Bacteriol. 173:3966-3976(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL B-18484 / PS811;
RA   Payne J.M., Sick A.J.;
RT   "Bacillus thuringiensis isolate active against lepidopteran pests, and

genes encoding novel lepidopteran-active toxins.";
RL   Patent number JP5189850; 23-FEB-1993.
CC   -!- FUNCTION: PROMOTES COLLOIDOSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC   EPITHELIAL CELLS OF MANY LEPIDOPTERAN LARVAE.
CC   -!- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC   SPORULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC   OF THE SPORE COAT.
CC   -!- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC   N-TERMINUS.
CC   -!- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC   -!- --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -!- --
DR   EMBL; M63897; AAA22348.1; -.
DR   EMBL; M73254; AAA22347.1; -.
DR   HSSP; P02965; 1CIY.
DR   InterPro; IPR001178; -.
DR   Pfam; PF00555; endotoxin; 1.
KW   Toxin; Sporulation.
SQ   SEQUENCE   1174 AA; 133621 MW; B51B9751D7791C61 CRC64;

Query Match   100.0%; Score 31; DB 1; Length 1174;
Best Local Similarity   100.0%; Pred. No. 1.1e+02;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
              ||||||
Db   50 PGVGVA 55

RESULT   5
ABL2_HUMAN
ID   ABL2_HUMAN   STANDARD;   PRT;   1182 AA.
AC   P42684;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   TYROSINE-PROTEIN KINASE ABL2 (EC 2.7.1.112) (TYROSINE KINASE ARG).
GN   ABL2 OR ARG OR ABL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90332670; PubMed=2198571;
RA   Kruh G.D., Perego R., Miki T., Aaronson S.A.;
RT   "The complete coding sequence of arg defines the Abelson subfamily of
RT   cytoplasmic tyrosine kinases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:5802-5806(1990).
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC   PROTEIN TYROSINE PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: IA AND IB (SHOWN HERE); ARE
CC   PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC   -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC   DOMAIN. BELONGS TO THE ABL SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   -!- --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial

```
CC   entries requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M35296; AAA35553.1; -.
DR   HSSP; P00520; 1ABO.
DR   MIM; 164690; -.
DR   InterPro; IPR000719; -.
DR   InterPro; IPR000980; -.
DR   InterPro; IPR001245; -.
DR   InterPro; IPR001452; -.
DR   Pfam; PF00017; SH2; 1.
DR   Pfam; PF00018; SH3; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   PRINTS; PR00401; SH2DOMAIN.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50001; SH2; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Transferase; Tyrosine-protein kinase; Proto-oncogene; ATP-binding;
KW   Phosphorylation; SH2 domain; SH3 domain; Alternative splicing.
FT   DOMAIN      107    167       SH3.
FT   DOMAIN      173    263       SH2.
FT   DOMAIN      288    539       PROTEIN KINASE.
FT   DOMAIN      561    564       POLY-SER.
FT   DOMAIN      658    660       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      732    739       POLY-GLY.
FT   DOMAIN      843   1055       PRO-RICH.
FT   DOMAIN      984    988       POLY-PRO.
FT   NP_BIND     294    302       ATP (BY SIMILARITY).
FT   BINDING     317    317       ATP (BY SIMILARITY).
FT   ACT_SITE    409    409       BY SIMILARITY.
FT   MOD_RES     439    439       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   VARSPLIC      1     73       MGQQVGRVGEAPGLQQPQPRGIRGSSAARPSGRRRDPAGRT
FT                                TETGTNFTTDHDHFASCVEDGFEGDKTGGSSP -> MVLGI
FT                                VLLPPNTYGRDQDTSLCCLCTEASESALPDLT (IN
FT                                ISOFORM IA).
SQ   SEQUENCE   1182 AA;  128343 MW;  ED93869BC2B14FAA CRC64;


Query Match             100.0%;  Score 31;  DB 1;  Length 1182;
Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PGVGVA 6
              ||||||
Db        877 PSVGVA 882


ID   UVRC_MYCTU     STANDARD;      PRT;   646 AA.
AC   P71689;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   EXCINUCLEASE ABC SUBUNIT C.
GN   UVRC OR RV1420 OR MTCY21B4.38.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC       CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC       PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S), ATTACHES TO THE
CC       UVRA-UVRB COMPLEX, DISPLACING UVRA, AND THE DAMAGED DNA STRAND IS
CC       NICKED ON BOTH SIDES OF THE DAMAGED SITE (BY SIMILARITY).
CC   -!- SUBUNIT: CONSISTS OF THREE SUBUNITS: UVRA, UVRB AND UVRC.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE UVRC FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z80108; CAB02168.1; -.
DR   TuberculList; Rv1420; -.
DR   InterPro; IPR000305; -.
DR   InterPro; IPR000445; -.
DR   Pfam; PF01541; Exci_endo_N; 1.
DR   Pfam; PF00633; HHH; 1.
KW   SOS response; Excision nuclease; DNA repair.
SQ   SEQUENCE   646 AA;  71614 MW;  CD7792F6C54E6709 CRC64;


Query Match              96.8%;  Score 30;  DB 1;  Length 646;
Best Local Similarity    83.3%;  Pred. No. 1e+02;
Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 PGVGVA 6
              ||:|||
Db        623 PGIGVA 628


RESULT 7
ID   DEF_MYCGE     STANDARD;      PRT;   216 AA.
AC   P47352;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   POLYPEPTIDE DEFORMYLASE (EC 3.5.1.31) (PDF) (FORMYLMETHIONINE
DE   DEFORMYLASE).
GN   DEF OR MG106.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
CC   -!- FUNCTION: REMOVES THE FORMYL GROUP FROM THE N-TERMINAL MET OF
CC       NEWLY SYNTHESIZED PROTEINS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N-FORMYL-L-METHIONINE + H(2)O = FORMATE +
CC       L-METHIONINE.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE POLYPEPTIDE DEFORMYLASE FAMILY.
```

CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U39691; AAC71324.1; ALT_INIT.
DR   HSSP; P27251; 1DEF.
DR   TIGR; MG106; -.
DR   InterPro; IPR000181; -.
DR   Pfam; PF01327; Pep_deformylase; 1.
KW   Protein biosynthesis; Hydrolase; Zinc.
FT   METAL   134   134   ZINC (BY SIMILARITY).
FT   METAL   178   178   ZINC (BY SIMILARITY).
FT   METAL   179   179   BY SIMILARITY.
FT   ACT_SITE   182   182   ZINC (BY SIMILARITY).
SQ   SEQUENCE   216 AA; 25171 MW; 52D22EBFA58BEFC5 CRC64;

Query Match   93.5%; Score 29; DB 1; Length 216;
Best Local Similarity   66.7%; Pred. No. 58;
Matches   4; Conservative   2; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
               |||||
Db   82 PGIGIA 87

RESULT 8
ID   DEF_MYCPN   STANDARD;   PRT;   216 AA.
AC   P75527;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   POLYPEPTIDE DEFORMYLASE (EC 3.5.1.31) (PDF) (FORMYLMETHIONINE
DE   DEFORMYLASE).
GN   DEF OR MPN246 OR MP587.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- FUNCTION: REMOVES THE FORMYL GROUP FROM THE N-TERMINAL MET OF
CC   NEWLY SYNTHESIZED PROTEINS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N-FORMYL-L-METHIONINE + H(2)O = FORMATE +
CC   L-METHIONINE.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE POLYPEPTIDE DEFORMYLASE FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AB000057; AAB96235.1; -.
DR   HSSP; P27251; 1DEF.
DR   InterPro; IPR000181; -.
DR   Pfam; PF01327; Pep_deformylase; 1.

KW   Protein biosynthesis; Hydrolase; Zinc.
FT   METAL   134   134   ZINC (BY SIMILARITY).
FT   METAL   178   178   ZINC (BY SIMILARITY).
FT   METAL   179   179   BY SIMILARITY.
FT   ACT_SITE   182   182   ZINC (BY SIMILARITY).
FT   METAL   182   182   ZINC (BY SIMILARITY).
SQ   SEQUENCE   216 AA; 24592 MW; 3C1F43048B505DF CRC64;

Query Match   93.5%; Score 29; DB 1; Length 216;
Best Local Similarity   66.7%; Pred. No. 58;
Matches   4; Conservative   2; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
               |:|:|
Db   82 PGIGIA 87

RESULT 9
COX6A_BOVIN
ID   COX6A_BOVIN   STANDARD;   PRT;   85 AA.
AC   P13182;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE VIA-LIVER (EC 1.9.3.1) (SSG).
GN   COX6A1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92070527; PubMed=1720401;
RA   Ewart G.D., Zhang Y.Z., Capaldi R.A.;
RT   "Switching of bovine cytochrome c oxidase subunit VIa isoforms in
RT   skeletal muscle during development.";
RL   FEBS Lett. 292:79-84(1991).
RN   [2]
RP   SEQUENCE OF 1-32.
RC   TISSUE=Liver;
RX   MEDLINE=89006097; PubMed=2844245;
RA   Yanamura W., Zhang Y.Z., Takamiya S., Capaldi R.A.;
RT   "Tissue-specific differences between heart and liver cytochrome c
RT   oxidase.";
RL   Biochemistry 27:4909-4914(1988).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE
CC   CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN
CC   MITOCHONDRIAL ELECTRON TRANSPORT.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC   4 FERRICYTOCHROME C.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VIA FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; M38520; AAA30437.1; -.
DR   PIR; B29968; B29968.
DR   HSSP; P07471; 1OCC.
DR   InterPro; IPR001349; -.
DR   Pfam; PF02046; COX6A; 1.
DR   PROSITE; PS01329; COX6A; 1.
KW   Oxidoreductase; Inner membrane; Mitochondrion.
FT   CONFLICT   19   19   L -> Y (IN REF. 2).
FT   CONFLICT   28   28   V -> L (IN REF. 2).
FT   CONFLICT   30   30   M -> T (IN REF. 2).

```
SQ   SEQUENCE   85 AA;   9507 MW;   9F493F6979E2E74B CRC64;

Query Match           90.3%;   Score 28;   DB 1;   Length 85;
Best Local Similarity 83.3%;   Pred. No. 38;
Matches   5;  Conservative   1;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 PGVGVA 6
         |||||
Db   24 PGVGVS 29

RESULT 10
YM74_MYCTU
ID   YM74_MYCTU   STANDARD;   PRT;   105 AA.
AC   Q50689;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   VERY HYPOTHETICAL 11.1 KDA PROTEIN RV2274C.
GN   RV2274C OR MTCY339.36.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton J., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   ----------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; Z77163; CAB00961.1; -.
DR   TubercuList; Rv2274c; -.
KW   Hypothetical protein.
SQ   SEQUENCE   105 AA;   11046 MW;   BBB3716A9ADDE26 CRC64;

Query Match           90.3%;   Score 28;   DB 1;   Length 105;
Best Local Similarity 83.3%;   Pred. No. 47;
Matches   5;  Conservative   1;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 PGVGVA 6
         |||||
Db   76 PGVGLA 81

RESULT 11
COXE_MOUSE
ID   COXE_MOUSE   STANDARD;   PRT;   111 AA.
AC   P43024;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE VIA-LIVER PRECURSOR (EC 1.9.3.1).
GN   COX6A1 OR COX6AL.
```

```
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=95178562; PubMed=7873616;
RA   Grossman L.I., Rosenthal N.H., Akamatsu K., Erickson R.P.;
RT   "Cloning, sequence analysis, and expression of a mouse cDNA encoding
RT   cytochrome c oxidase subunit via liver isoform.";
RL   Biochim. Biophys. Acta 1260:361-364(1995).
RN   [2]
RP   SEQUENCE OF 2-111 FROM N.A.
RC   STRAIN=BALB/C;
RA   Newton D., Bowman L.H.;
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE
CC       CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN
CC       MITOCHONDRIAL ELECTRON TRANSPORT.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VIA FAMILY.
CC   ----------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------------
DR   EMBL; L06465; AAA53065.1; ALT_INIT.
DR   EMBL; U08440; AAA17836.1; ALT_INIT.
DR   HSSP; P07471; 1OCC.
DR   SWISS-2DPAGE; P43024; MOUSE.
DR   MGD; MGI:103099; Cox6a1.
DR   InterPro; IPR001349; -.
DR   Pfam; PF02046; COX6A; 1.
DR   PROSITE; PS01329; COX6A; 1.
KW   Oxidoreductase; Inner membrane; Mitochondrion; Transit peptide.
FT   TRANSIT       1     26       MITOCHONDRION (BY SIMILARITY).
FT   CHAIN        27    111       CYTOCHROME C OXIDASE POLYPEPTIDE VIA-
FT                                LIVER.
SQ   SEQUENCE   111 AA;   12352 MW;   76D5EDA5F79EC5E7 CRC64;

Query Match           90.3%;   Score 28;   DB 1;   Length 111;
Best Local Similarity 83.3%;   Pred. No. 49;
Matches   5;  Conservative   1;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 PGVGVA 6
         |||||
Db   50 PGVGVS 55

RESULT 12
COXE_RAT
ID   COXE_RAT   STANDARD;   PRT;   111 AA.
AC   P10818;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE VIA-LIVER PRECURSOR (EC 1.9.3.1).
GN   COX6A1 OR COX6AL.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISCHER; TISSUE=Liver;
```

```
RX   MEDLINE=94192996; PubMed=8144025;
RA   Mell O.C., Seibel P., Kadenbach B.;
RT   "Structural organisation of the rat genes encoding liver- and
RT   heart-type of cytochrome c oxidase subunit VIa and a pseudogene
RT   related to the COXVIa-L cDNA.";
RL   Gene 140:179-186(1994).
RN   [2]
RP   SEQUENCE OF 27-111 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=89052650; PubMed=2461293;
RA   Schlerf A., Droste M., Winter M., Kadenbach B.;
RT   "Characterization of two different genes (cDNA) for cytochrome c
RT   oxidase subunit VIa from heart and liver of the rat.";
RL   EMBO J. 7:2387-2391(1988).
RN   [3]
RP   SEQUENCE OF 27-48.
RC   TISSUE=Liver, and Brown adipose tissue;
RX   MEDLINE=90122894; PubMed=2153407;
RA   Kadenbach B., Stroh A., Becker A., Eckerson C., Lottspeich F.;
RT   "Tissue- and species-specific expression of cytochrome c oxidase
RT   isozymes in vertebrates.";
RL   Biochim. Biophys. Acta 1015:368-372(1990).
RN   [4]
RP   SEQUENCE OF 27-49.
RC   STRAIN=WISTAR; TISSUE=Liver;
RX   MEDLINE=95324529; PubMed=7601105;
RA   Schaeeger H., Noack H., Halangk W., Brandt U., von Jagow G.;
RT   "Cytochrome-c oxidase in developing rat heart. Enzymic properties and
RT   amino-terminal sequences suggest identity of the fetal heart and the
RT   adult liver isoform.";
RL   Eur. J. Biochem. 230:235-241(1995).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE
CC       CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN
CC       MITOCHONDRIAL ELECTRON TRANSPORT.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VIA FAMILY.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; X72757; CAA51286.1; ALT_SEQ.
DR   EMBL; X12553; CAA31067.1; -.
DR   PIR; S01156; S01156.
DR   HSSP; P07471; 1OCC.
DR   InterPro; IPR001349; -.
DR   Pfam; PF02046; COX6A; 1.
DR   PROSITE; PS01329; COX6A; 1.
KW   Oxidoreductase; Inner membrane; Mitochondrion; Transit peptide.
FT   TRANSIT       1     26       MITOCHONDRION.
FT   CHAIN        27    111       CYTOCHROME C OXIDASE POLYPEPTIDE VIA-
FT                                LIVER.
SQ   SEQUENCE   111 AA;  12301 MW;  71EA06B23ED754E6D CRC64;

Query Match             90.3%;  Score 28;  DB 1;  Length 111;
Best Local Similarity   83.3%;  Pred. No. 49;
Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 PGVGVA 6
            |||||
Db     50 PGVGVS 55

RESULT 13
DEF2_BACSU
ID   DEF2_BACSU    STANDARD;      PRT;     184 AA.
```

```
AC   Q45495;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   POLYPEPTIDE DEFORMYLASE 2 (EC 3.5.1.31) (PDF 2) (FORMYLMETHIONINE
DE   DEFORMYLASE 2).
GN   YKRB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Caldwell R.M., Ferrari E.;
RT   "Sequence analysis of the moaA-ampS region of the Bacillus subtilis
RT   chromosome.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REMOVES THE FORMYL GROUP FROM THE N-TERMINAL MET OF
CC       NEWLY SYNTHESIZED PROTEINS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N-FORMYL-L-METHIONINE + H(2)O = FORMATE +
CC       L-METHIONINE.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE POLYPEPTIDE DEFORMYLASE FAMILY.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AF012295; AAC24930.1; -.
DR   EMBL; Z99111; CAB13329.1; -.
DR   Subtilist; BG11815; YkrB.
DR   InterPro; IPR000181; -.
DR   Pfam; PF01327; Pep_deformylase; 1.
KW   Protein biosynthesis; Hydrolase; Zinc.
FT   METAL       110    110       ZINC (BY SIMILARITY).
FT   METAL       153    153       ZINC (BY SIMILARITY).
FT   ACT_SITE    154    154       BY SIMILARITY.
FT   METAL       157    157       ZINC (BY SIMILARITY).
SQ   SEQUENCE   184 AA;  20655 MW;  8641BF193266C38 CRC64;

Query Match             90.3%;  Score 28;  DB 1;  Length 184;
Best Local Similarity   83.3%;  Pred. No. 79;
Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PGVGVA 6
            ||||
Db     57 PGVGLA 62

RESULT 14
Y152_HUMAN
ID   Y152_HUMAN     STANDARD;      PRT;     292 AA.
AC   Q14165;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN KIAA0152.
GN   KIAA0152.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96127530; PubMed=8590280;
RA   Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
```

RT   "Prediction of the coding sequences of unidentified human genes. IV.
RT   The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RL   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 2:167-174(1995).
CC   -!- SIMILARITY: TO C.ELEGANS F44E2.4.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; D63486; BAA09773.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM   270   290   POTENTIAL.
FT   DOMAIN   14   22   POLY-LEU.
FT   DOMAIN   231   238   POLY-GLU.
SQ   SEQUENCE   292 AA;   32234 MW;   448D6773A5A1A8F09 CRC64;

Query Match   90.3%;   Score 28;   DB 1;   Length 292;
Best Local Similarity   83.3%;   Pred. No. 1.2e+02;
Matches   5;   Conservative   1;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1   PGVGVA   6
           ||:|||
Db   29   PGLGVA   34

RESULT 15
ID   IDHP_CANTR   STANDARD;   PRT;   430 AA.
AC   O13285;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ISOCITRATE DEHYDROGENASE [NADP], MITOCHONDRIAL PRECURSOR (EC 1.1.1.42)
DE   (OXALOSUCCINATE DECARBOXYLASE) (IDH) (NADP+-SPECIFIC ICDH) (IDP).
GN   IDP1.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5482;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PK233;
RX   MEDLINE=98004564; PubMed=9325427;
RA   Imajo T., Kawachi H., Atomi H., Sanuki S., Yamamoto S., Ueda M.,
RA   Tanaka A.;
RT   "Immunochemically distinct NADP-linked isocitrate dehydrogenase
RT   isozymes in mitochondria and peroxisomes of Candida tropicalis.";
RL   Arch. Microbiol. 168:389-395(1997).
CC   -!- FUNCTION: MITOCHONDRIAL IDP1 MAY REGULATE FLUX THROUGH THE
CC   TRICARBOXYLIC ACID CYCLE AND RESPIRATION. ITS PROBABLY CRITICAL
CC   FUNCTION IS THE PRODUCTION OF NADPH.
CC   -!- CATALYTIC ACTIVITY: ISOCITRATE + NADP(+) = 2-OXOGLUTARATE +
CC   CO(2) + NADPH.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC   -!- SIMILARITY: BELONGS TO THE ISOCITRATE AND ISOPROPYLMALATE
CC   DEHYDROGENASES FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AB004556; BAA22945.1; -.

DR   InterPro; IPR001804; -.
DR   Pfam; PF001804; isocdh; 1.
DR   PROSITE; PS00470; IDH_IMDH; 1.
KW   Oxidoreductase; NADP; Glyoxylate bypass; Tricarboxylic acid cycle;
KW   Transit peptide; Mitochondrion.
FT   TRANSIT   1   27   MITOCHONDRION (BY SIMILARITY).
FT   CHAIN   28   430   ISOCITRATE DEHYDROGENASE [NADP].
FT   ACT_SITE   120   120   BINDING TO ISOCITRATE (BY SIMILARITY).
SQ   SEQUENCE   430 AA;   48009 MW;   205A31949670CCEB CRC64;

Query Match   90.3%;   Score 28;   DB 1;   Length 430;
Best Local Similarity   83.3%;   Pred. No. 1.7e+02;
Matches   5;   Conservative   1;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1   PGVGVA   6
           ||:|||
Db   201   PGLGVA   205

Search completed: April 24, 2001, 16:42:49
Job time: 455 sec

```
                 GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

QM protein - protein search, using sw model

Run on:      April 24, 2001, 16:40:32 ; Search time 125.5 Seconds
                                (without alignments)
                                5.604 Million cell updates/sec

Title:       US-09-340-736-5
Perfect score: 31
Sequence:    1 PGVGVA 6

Scoring table: BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters:      374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL_15:*
             1: sp_archea:*
             2: sp_bacteria:*
             3: sp_fungi:*
             4: sp_human:*
             5: sp_invertebrate:*
             6: sp_mammal:*
             7: sp_mhc:*
             8: sp_organelle:*
             9: sp_phage:*
            10: sp_plant:*
            11: sp_rodent:*
            12: sp_unclassified:*
            13: sp_vertebrate:*
            14: sp_virus:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	100.0	76	6	Q28100	Q28100 bos taurus
2	31	100.0	258	4	Q90MF5	Q9umf5 homo sapien
3	31	100.0	458	5	Q9VYC4	Q9vyc4 drosophila
4	31	100.0	510	5	Q9VL73	Q9vl73 drosophila
5	31	100.0	540	5	Q9VRW0	Q9vrw0 drosophila
6	31	100.0	602	4	O15337	O15337 homo sapien
7	31	100.0	635	4	O15336	O15336 homo sapien
8	31	100.0	650	6	Q28099	Q28099 bos taurus
9	31	100.0	666	6	Q28096	Q28096 bos taurus
10	31	100.0	679	6	Q28097	Q28097 bos taurus
11	31	100.0	687	4	Q14235	Q14235 homo sapien
12	31	100.0	707	6	Q28098	Q28098 bos taurus
13	31	100.0	722	2	Q9XC22	Q9xc22 mycoplasma
14	31	100.0	724	4	Q14233	Q14233 homo sapien
15	31	100.0	747	5	Q9VWN1	Q9vwn1 drosophila
16	31	100.0	757	4	Q14234	Q14234 homo sapien
17	31	100.0	863	5	Q9V329	Q9v329 drosophila
18	31	100.0	907	5	Q26675	Q26675 theileria a
19	31	100.0	1174	2	O45749	O45749 bacillus th

20	30	96.8	92	5	Q9NYI0	Q9nyi0 leishmania
21	30	96.8	244	5	Q9NEF1	Q9nefi leishmania
22	30	96.8	358	10	Q9M333	Q9m333 arabidopsis
23	30	96.8	456	2	Q9ZBA6	Q9zba6 streptomyce
24	29	93.5	745	5	Q9U195	Q9u195 leishmania
25	29	93.5	204	2	Q9XX63	Q9xx63 mycoplasma
26	29	93.5	317	2	Q92533	Q92533 streptomyce
27	29	93.5	390	10	Q92828	Q92828 chlamydia p
28	28	90.3	617	10	Q9LXY2	Q9lxy2 arabidopsis
29	28	90.3	86	4	O95101	O95101 homo sapien
30	28	90.3	95	5	O77095	O77095 caenorhabdi
31	28	90.3	97	5	Q9VC12	Q9vc12 drosophila
32	28	90.3	109	6	Q9TT7	Q9tt7 oryctolagus
33	28	90.3	182	2	Q9K9I9	Q9k9i9 bacillus ha
34	28	90.3	218	2	Q9RD27	Q9rd27 streptomyce
35	28	90.3	253	5	Q9VBI2	Q9vbi2 drosophila
36	28	90.3	259	10	Q9XI30	Q9xi30 arabidopsis
37	28	90.3	280	10	O82173	O82173 arabidopsis
38	28	90.3	312	5	Q9UZA0	Q9uza0 caenorhabdi
39	28	90.3	322	5	O45024	O45024 branchiosto
40	28	90.3	354	5	Q9VVU8	Q9vvu8 drosophila
41	28	90.3	368	2	Q44244	Q44244 amycolatops
42	28	90.3	497	5	Q9LOI4	Q9loi4 streptomyce
43	28	90.3	501	10	Q9SZ01	Q9sz0l arabidopsis
44	28	90.3	541	2	Q9RN71	Q9rn71 streptomyce
45	28	90.3	543	10	O22188	O22188 arabidopsis

ALIGNMENTS

```
RESULT   1
Q28100
ID   Q28100           PRELIMINARY;      PRT;       76 AA.
AC   Q28100;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85059254; PubMed=6150137;
RA   Rosenbloom J.;
RT   "Elastin: relation of protein and gene structure to disease.";
RL   Lab. Invest. 51:605-623(1984).
DR   EMBL; M31891; AAA96416.1; -.
DR   EMBL; M31893; AAA96416.1; JOINED.
DR   EMBL; M31892; AAA96416.1; JOINED.
FT   NON_TER       1      1
SQ   SEQUENCE   76 AA;    6619 MW;   E68337 9DAB87B202 CRC64;

Query Match        100.0%;  Score 31;  DB 6;  Length 76;
Best Local Similarity  100.0%;  Pred. No. 36;
Matches   6;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 PGVGVA 6
            ||||||
Db     23 PGVGVA 28

RESULT   2
Q90MF5
ID   Q90MF5           PRELIMINARY;      PRT;      258 AA.
AC   Q90MF5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
```

DE   ELASTIN (FRAGMENT).
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96411691; PubMed=8812450;
RA   Osborne L.R., Martindale D.W., Scherer S.W., Shi X.-M., Huizenga J.,
RA   Heng H.H.Q., Costa T., Pober B., Lew L., Brinkman J., Rommens J.,
RA   Koop B.F., Tsui L.-C.;
RT   "Identification of genes from a 500-kb region at 7q11.23 that is
RT   commonly deleted in Williams syndrome patients.";
RL   Genomics 36:328-336(1996).
DR   EMBL; D63721; AAC13884.1; -.
DR   INTERPRO; IPR001451; -.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
FT   NON_TER   1   1
SQ   SEQUENCE   258 AA; 21990 MW; C39BF7298DCE05D2 CRC64;

Query Match   100.0%; Score 31; DB 4; Length 258;
Best Local Similarity   100.0%; Pred. No. 1.3e+02;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
              ||||||
Db   9 PGVGVA 14

RESULT 3
ID   Q9VYC4
AC   Q9VYC4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG15753 PROTEIN.
GN   CG15753.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA   Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003492; AAF48279.1; -.
DR   FLYBASE; FBgn0030491; CG15753.
SQ   SEQUENCE   458 AA; 50794 MW; 323B5D107FB250F3 CRC64;

Query Match   100.0%; Score 31; DB 5; Length 458;
Best Local Similarity   100.0%; Pred. No. 2.4e+02;
Matches   6; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 PGVGVA 6
              ||||||
Db   143 PGVGVA 148

RESULT 4
ID   Q9VL73
AC   Q9VL73;   PRELIMINARY;   PRT;   510 AA.
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG13124
GN   CG13124.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Venter R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003626; AAF52823.1; -.
DR   FLYBASE; FBgn0032156; CG13124.
SQ   SEQUENCE 510 AA; 55866 MW; D6F2B3B58D15386A CRC64;

Query Match   100.0%; Score 31; DB 5; Length 510;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 PGVGVA 6
         ||||||
Db   142 PGVGVA 147

RESULT 5
Q9VRM0
ID   Q9VRM0   PRELIMINARY;   PRT;   540 AA.
AC   Q9VRM0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG10626 PROTEIN.
GN   CG10626.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003566; AAF50775.1; -.
DR   FLYBASE; FBgn0035610; CG10626.
DR   INTERPRO; IPR000276; -.
DR   INTERPRO; IPR000611; -.
DR   INTERPRO; IPR001556; -.
DR   PFAM; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PRINTS; PR00358; BOMBESINR.
DR   PRINTS; PR01012; NRPEPTIDEYR.
DR   PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
SQ   SEQUENCE 540 AA; 60612 MW; 2D22D5E1BDFD47D5 CRC64;

Query Match   100.0%; Score 31; DB 5; Length 540;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 PGVGVA 6
         ||||||
Db   435 PGVGVA 440

RESULT 6
O15337
ID   O15337   PRELIMINARY;   PRT;   602 AA.
AC   O15337;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Li D.Y., Toland A.B., Boak B.B., Atkinson D.L., Ensing G.J.,
RA   Morris C.A., Keating M.T.;
RL   Hum. Mol. Genet. 0:0-0(1997).
DR   EMBL; U93037; AAB85620.1; -.
DR   EMBL; U93034; AAB85620.1; JOINED.
DR   EMBL; U93035; AAB85620.1; JOINED.
DR   EMBL; U93036; AAB85620.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   INTERPRO; IPR001807; -.
DR   INTERPRO; IPR002185; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00569; DOPAMINED4R.
DR   PRINTS; PR00762; CLCHANNEL.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
FT   NON_TER   1   1
FT   NON_TER   602   602
SQ   SEQUENCE 602 AA; 51807 MW; 5JB5B9A71EF04807 CRC64;

Query Match   100.0%; Score 31; DB 4; Length 602;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 PGVGVA 6
         ||||||
Db   482 PGVGVA 487

RESULT 7
Q15336
```
ID   Q15336      PRELIMINARY;      PRT;   635 AA.
AC   Q15336;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Li D.Y., Toland A.E., Boak B.B., Atkinson D.L., Ensing G.J.,
RA   Morris C.A., Keating M.T.;
RL   Hum. Mol. Genet. 0:0-0(1997).
DR   EMBL; U93037; AAB55621.1; -.
DR   EMBL; U93034; AAB55621.1; JOINED.
DR   EMBL; U93035; AAB55621.1; JOINED.
DR   EMBL; U93036; AAB55621.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   INTERPRO; IPR001807; -.
DR   INTERPRO; IPR002185; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00559; DOPAMINE4R.
DR   PRINTS; PR00762; CLCHANNEL.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
FT   NON_TER       1      1
FT   NON_TER     635    635
SQ   SEQUENCE   635 AA;   55279 MW;   72950C364127B2A4 CRC64;

Query Match             100.0%;  Score 31;  DB 4;  Length 635;
Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
Matches    6;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 PGVGVA 6
            ||||||
Db      482 PGVGVA 487
```

RESULT 8
Q28099
```
ID   Q28099      PRELIMINARY;      PRT;   650 AA.
AC   Q28099;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN-CBEl (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992576;
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J., Yoon K.;
RT   "Structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Cicila G., Yoon K., Rosenbloom J.;
RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30499.1; -.
DR   EMBL; M11422; AAA30499.1; JOINED.
DR   EMBL; M19366; AAA30499.1; JOINED.
DR   EMBL; M19368; AAA30499.1; JOINED.
DR   EMBL; M19369; AAA30499.1; JOINED.
DR   EMBL; M19370; AAA30499.1; JOINED.
DR   EMBL; M19371; AAA30499.1; JOINED.
DR   EMBL; M22771; AAA30499.1; JOINED.
DR   EMBL; M22772; AAA30499.1; JOINED.
DR   EMBL; M22773; AAA30499.1; JOINED.
DR   EMBL; M22774; AAA30499.1; JOINED.
DR   EMBL; M22775; AAA30499.1; JOINED.
DR   EMBL; M29988; AAA30499.1; JOINED.
DR   EMBL; M30100; AAA30499.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
FT   NON_TER       1      1
SQ   SEQUENCE   650 AA;   55373 MW;   CD21ABB3B9076AD7 CRC64;

Query Match             100.0%;  Score 31;  DB 6;  Length 650;
Best Local Similarity   100.0%;  Pred. No. 3.5e+02;
Matches    6;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 PGVGVA 6
            ||||||
Db      412 PGVGVA 417
```

RESULT 9
Q28096
```
ID   Q28096      PRELIMINARY;      PRT;   666 AA.
AC   Q28096;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN-CBEL3 (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992576;
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J., Yoon K.;
RT   "Structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Cicila G., Yoon K., Rosenbloom J.;
RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30501.1; -.
DR   EMBL; M11422; AAA30501.1; JOINED.
DR   EMBL; M19366; AAA30501.1; JOINED.
DR   EMBL; M19367; AAA30501.1; JOINED.
DR   EMBL; M19368; AAA30501.1; JOINED.
DR   EMBL; M19369; AAA30501.1; JOINED.
DR   EMBL; M19370; AAA30501.1; JOINED.
DR   EMBL; M19371; AAA30501.1; JOINED.
DR   EMBL; M22771; AAA30501.1; JOINED.
DR   EMBL; M22772; AAA30501.1; JOINED.
DR   EMBL; M22773; AAA30501.1; JOINED.
DR   EMBL; M22774; AAA30501.1; JOINED.
DR   EMBL; M22775; AAA30501.1; JOINED.
DR   EMBL; M29988; AAA30501.1; JOINED.
```

```
DR   INTERPRO; IPR000104; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
FT   NON_TER       1      1
SQ   SEQUENCE   666 AA;  56435 MW;  BC35E62632BE1B71 CRC64;

Query Match             100.0%;  Score 31;  DB 6;  Length 666;
Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PGVGVA 6
              ||||||
Db        420 PGVGVA 425


RESULT 10
Q28097
ID   Q28097      PRELIMINARY;       PRT;   579 AA.
AC   Q28097;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN-CBEL2 (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992576;
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J., Yoon K.;
RT   "Structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Cicila G., Yoon K., Rosenbloom J.;
RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30500.1; -.
DR   EMBL; M11422; AAA30500.1; JOINED.
DR   EMBL; M19366; AAA30500.1; JOINED.
DR   EMBL; M19367; AAA30500.1; JOINED.
DR   EMBL; M19368; AAA30500.1; JOINED.
DR   EMBL; M19369; AAA30500.1; JOINED.
DR   EMBL; M19370; AAA30500.1; JOINED.
DR   EMBL; M19371; AAA30500.1; JOINED.
DR   EMBL; M22771; AAA30500.1; JOINED.
DR   EMBL; M22772; AAA30500.1; JOINED.
DR   EMBL; M22773; AAA30500.1; JOINED.
DR   EMBL; M22774; AAA30500.1; JOINED.
DR   EMBL; M22988; AAA30500.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   INTERPRO; IPR001459; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00959; MEVGALKINASE.
FT   NON_TER       1      1
SQ   SEQUENCE   679 AA;  57652 MW;  EB3C019E3BD7618D CRC64;

Query Match             100.0%;  Score 31;  DB 6;  Length 679;
Best Local Similarity   100.0%;  Pred. No. 3.7e+02;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PGVGVA 6
              ||||||
Db        446 PGVGVA 451


RESULT 11
Q14235
ID   Q14235      PRELIMINARY;       PRT;   687 AA.
AC   Q14235;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN.
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87274906; PubMed=3038460;
RA   Indik Z., Yoon K., Morrow S.D., Cicila G., Rosenbloom J.,
RA   Rosenbloom J., Ornstein-Goldstein N.;
RT   "Structure of the 3' region of the human elastin gene: great abundance
RT   of Alu repetitive sequences and few coding sequences.";
RL   Connect. Tissue Res. 16:197-211(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87289666; PubMed=3039501;
RA   Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT   "Alternative splicing of human elastin mRNA indicated by sequence
RT   analysis of cloned genomic and complementary DNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
DR   EMBL; M17282; AAC98393.1; -.
DR   EMBL; M16983; AAC98393.1; JOINED.
DR   EMBL; M17265; AAC98393.1; JOINED.
DR   EMBL; M17266; AAC98393.1; JOINED.
DR   EMBL; M17267; AAC98393.1; JOINED.
DR   EMBL; M17268; AAC98393.1; JOINED.
DR   EMBL; M17271; AAC98393.1; JOINED.
DR   EMBL; M17272; AAC98393.1; JOINED.
DR   EMBL; M17273; AAC98393.1; JOINED.
DR   EMBL; M17275; AAC98393.1; JOINED.
DR   EMBL; M17276; AAC98393.1; JOINED.
DR   EMBL; M17277; AAC98393.1; JOINED.
DR   EMBL; M17278; AAC98393.1; JOINED.
DR   EMBL; M17279; AAC98393.1; JOINED.
DR   EMBL; M17281; AAC98393.1; JOINED.
DR   INTERPRO; IPR001179; -.
DR   INTERPRO; IPR001451; -.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR   PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
SQ   SEQUENCE   687 AA;  59579 MW;  9D5AC3C4D9F9B98E CRC64;

Query Match             100.0%;  Score 31;  DB 4;  Length 687;
Best Local Similarity   100.0%;  Pred. No. 3.7e+02;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 PGVGVA 6
              ||||||
Db        455 PGVGVA 461


RESULT 12
Q28098
ID   Q28098      PRELIMINARY;       PRT;   707 AA.
AC   Q28098;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN (FRAGMENT).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
```

```
RN   [1]
RP   SEQUENCE OF 17-35 FROM N.A.
RX   MEDLINE=85280426; PubMed=2992576;
RA   Cicila G., May M., Ornstein-Goldstein N., Indik Z., Morrow S.,
RA   Yeh H.S., Rosenbloom J., Boyd C., Rosenbloom J., Yoon K.;
RT   "Structure of the 3' portion of the bovine elastin gene.";
RL   Biochemistry 24:3075-3080(1985).
RN   [2]
RP   SEQUENCE OF 1-35 FROM N.A.
RX   MEDLINE=88028442; PubMed=3665402;
RA   Yeh H., Ornstein-Goldstein N., Indik Z., Sheppard P., Anderson N.,
RA   Rosenbloom J.C. Cicila G., Yoon K., Rosenbloom J.;
RT   "Sequence variation of bovine elastin mRNA due to alternative
RT   splicing.";
RL   Coll. Relat. Res. 7:235-247(1987).
DR   EMBL; M19372; AAA30498.1; -.
DR   EMBL; M11422; AAA30498.1; JOINED.
DR   EMBL; M19366; AAA30498.1; JOINED.
DR   EMBL; M19367; AAA30498.1; JOINED.
DR   EMBL; M19368; AAA30498.1; JOINED.
DR   EMBL; M19369; AAA30498.1; JOINED.
DR   EMBL; M19370; AAA30498.1; JOINED.
DR   EMBL; M19371; AAA30498.1; JOINED.
DR   EMBL; M22771; AAA30498.1; JOINED.
DR   EMBL; M22772; AAA30498.1; JOINED.
DR   EMBL; M22773; AAA30498.1; JOINED.
DR   EMBL; M22774; AAA30498.1; JOINED.
DR   EMBL; M22775; AAA30498.1; JOINED.
DR   EMBL; M22988; AAA30498.1; JOINED.
DR   EMBL; M23010; AAA30498.1; JOINED.
DR   INTERPRO; IPR000104; -.
DR   INTERPRO; IPR001459; -.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00959; NEVGALKINASE.
FT   NON_TER    1      1
SQ   SEQUENCE   707 AA;   60346 MW;  FDFD569BAB34CE33 CRC64;

Query Match             100.0%;  Score 31;  DB 6;  Length 707;
Best Local Similarity   100.0%;  Pred. No. 3.8e+02;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1  PGVGVA  6
          ||||||
Db   446  PGVGVA  451


RESULT 13
Q9XC22
Q9XC22      PRELIMINARY;      PRT;   722 AA.
            01-NOV-1999 (TrEMBLrel. 12, Created)
            01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
            01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE ALPHA CHAIN (EC 1.17.4.1).
GN   NRDE.
OS   Mycoplasma gallisepticum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2096;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=a5969;
RA   Skamrov A., Goldman M., Feoktistova E., Beabealashvilli R.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS.
CC   -!- CATALYTIC ACTIVITY: 2' DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC       THIOREDOXIN = H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC       THIOREDOXIN.
CC   -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
DR   EMBL; AF152114; AAD45273.1; -.
DR   INTERPRO; IPR000788; -.
DR   PFAM; PF00317; ribonucleo_red; 1.
```

```
DR   PRINTS; PR01183; RIBORDTASDM.
DR   PROSITE; PS00089; RIBORED_LARGE; 1.
KW   Oxidoreductase; DNA replication.
SQ   SEQUENCE   722 AA;   81837 MW;  C15B17A8F976B330 CRC64;

Query Match             100.0%;  Score 31;  DB 2;  Length 722;
Best Local Similarity   100.0%;  Pred. No. 3.9e+02;
Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1  PGVGVA  6
          ||||||
Db   245  PGVGVA  250


RESULT 14
Q14233
Q14233      PRELIMINARY;      PRT;   724 AA.
AC   Q14233; Q14238;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELASTIN.
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87274906; PubMed=3038460;
RA   Indik Z., Yoon K., Morrow S.D., Cicila G., Rosenbloom J.,
RA   Rosenbloom J., Ornstein-Goldstein N.;
RT   "Structure of the 3' region of the human elastin gene: great abundance
RT   of Alu repetitive sequences and few coding sequences.";
RL   Connect. Tissue Res. 16:197-211(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87289668; PubMed=3039501;
RA   Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT   "Alternative splicing of human elastin mRNA indicated by sequence
RT   analysis of cloned genomic and complementary DNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
RN   [3]
RP   SEQUENCE OF 164-724 FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE=88156138; PubMed=2831431;
RA   Fazio M.J., Olsen D.R., Kuivaniemi H., Chu M.L., Davidson J.M.,
RA   Rosenbloom J., Uitto J.;
RT   "Isolation and characterization of human elastin cDNAs, and age-
RT   associated variation in elastin gene expression in cultured skin
RT   fibroblasts.";
RL   Lab. Invest. 58:270-277(1988).
DR   EMBL; M17282; AAC98394.1; JOINED.
DR   EMBL; M16983; AAC98394.1; JOINED.
DR   EMBL; M17265; AAC98394.1; JOINED.
DR   EMBL; M17266; AAC98394.1; JOINED.
DR   EMBL; M17267; AAC98394.1; JOINED.
DR   EMBL; M17268; AAC98394.1; JOINED.
DR   EMBL; M17270; AAC98394.1; JOINED.
DR   EMBL; M17271; AAC98394.1; JOINED.
DR   EMBL; M17272; AAC98394.1; JOINED.
DR   EMBL; M17273; AAC98394.1; JOINED.
DR   EMBL; M17275; AAC98394.1; JOINED.
DR   EMBL; M17276; AAC98394.1; JOINED.
DR   EMBL; M17277; AAC98394.1; JOINED.
DR   EMBL; M17278; AAC98394.1; JOINED.
DR   EMBL; M17279; AAC98394.1; JOINED.
DR   EMBL; M17280; AAC98394.1; JOINED.
DR   EMBL; M17281; AAC98394.1; JOINED.
DR   EMBL; M24782; AAA53190.1; JOINED.
DR   INTERPRO; IPR001179; -.
```

DR INTERPRO; IPR001451; -.
DR PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
DR PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
SQ SEQUENCE 724 AA; 62664 MW; 6115568228EDF3D CRC64;

Query Match   100.0%; Score 31; DB 4; Length 724;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 PGVGVA 6
     ||||||
Db 475 PGVGVA 480

RESULT 15
Q9VVN1
Q9VVN1   PRELIMINARY;   PRT; 747 AA.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG12527 PROTEIN.
GN CG12527.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003510; AAF48907.1; -.
DR FLYBASE; FBgn0030977; CG12527.

SQ SEQUENCE 747 AA; 76047 MW; 3008D28A92C4EEF1 CRC64;

Query Match   100.0%; Score 31; DB 5; Length 747;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 PGVGVA 6
     ||||||
Db 680 PGVGVA 685

Search completed: April 24, 2001, 16:40:35
Job time: 425 sec

us-09-340-736-5.rspt

```
              GenCore version 4.5

         Copyright (c) 1993 - 2000  Compugen Ltd.


OM protein - protein search, using sw model

Run on:      April 24, 2001, 16:38:19 ; Search time 115.25 seconds
                                (without alignments)
                                2.480 Million cell updates/sec

Title:        US-09-340-736-6
Perfect score: 27
Sequence:     1 VPGVG 5

Scoring table: BLOSUM62
              Gapop 10.0 , Gapext 0.5

Searched:     390729 seqs, 57163235 residues


Total number of hits satisfying chosen parameters:    390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    A_Geneseq_0401:*
    1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
    2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
    3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
    4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
    5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
    6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
    7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
    8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
    9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
   10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
   11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
   12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
   13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
   14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
   15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
   16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
   17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
   18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
   19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
   20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
   21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
   22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

                      SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

		%	Query				
Result No.	Score	Match	Length	DB	ID	Description	
----	-----	-----	------	--	--	-----------	
1	27	100.0	5	7	P61199	Repeating unit.	
2	27	100.0	5	13	R29149	Pentapeptide repea	
3	27	100.0	5	13	R29145	Bioelastic pentape	
4	27	100.0	5	16	R65228	Fibrinogen/beta-ca	
5	27	100.0	5	16	R80250	Elastin-like repeat	
6	27	100.0	5	16	R80307	Elastin cross-link	
7	27	100.0	5	18	W22714	Bioelastomeric rep	
8	27	100.0	5	18	W26332	Elastin repeat uni	
9	27	100.0	5	18	W18261	Transglutaminase c	
10	27	100.0	5	18	W12301	Monomeric unit for	
11	27	100.0	5	19	W49701	Elastin repeat mot	

12	27	100.0	5	19	W53506	Peptide reapeat un
13	27	100.0	5	19	W47348	Elastin repeat uni
14	27	100.0	5	20	Y55877	Apolipoprotein fra
15	27	100.0	5	20	Y31686	pentapeptide used
16	27	100.0	5	20	Y33102	Elastin derived re
17	27	100.0	5	21	B19211	Elastin fragment u
18	27	100.0	5	21	Y80334	Repeat elastin pen
19	27	100.0	5	21	Y78246	Repeating unit. SEQ
20	27	100.0	5	21	Y78265	Antigen peptide se
21	27	100.0	6	22	B63972	Elastin repeating
22	27	100.0	6	19	W47335	Inhibitor of LDL b
23	27	100.0	6	20	Y55874	Apolipoprotein fra
24	27	100.0	6	20	Y33099	Elastin derived pe
25	27	100.0	8	20	Y06866	Short elastin pept
26	27	100.0	8	20	Y06867	Short elastin pept
27	27	100.0	13	10	P91307	Sequence of beta-t
28	27	100.0	15	18	W34883	Bioelastomeric pol
29	27	100.0	15	20	Y55862	Apolipoprotein fra
30	27	100.0	16	19	W47350	Inhibitor of LDL b
31	27	100.0	16	20	Y55879	Apolipoprotein fra
32	27	100.0	16	20	Y33104	Elastin derived re
33	27	100.0	18	19	W44902	Bovine "polyprolin
34	27	100.0	18	20	Y06871	Short elastin pept
35	27	100.0	19	20	Y06872	Short elastin pept
36	27	100.0	20	13	R26383	Sequence of peptid
37	27	100.0	20	14	R40000	Scintigraph imagin
38	27	100.0	20	17	W11073	Atherosclerotic pl
39	27	100.0	20	20	W31123	Atherosclerotic pl
40	27	100.0	20	20	Y55861	Apolipoprotein fra
41	27	100.0	20	20	Y06873	Short elastin pept
42	27	100.0	21	12	R15135	Vascular injury af
43	27	100.0	21	14	R42543	Leukocyte-binding
44	27	100.0	21	19	W47336	Inhibitor of LDL b
45	27	100.0	21	20	Y55875	Apolipoprotein fra

```
                                                    ALIGNMENTS

RESULT      1
P61199
ID   P61199 standard; protein; 5 AA.
XX
AC   P61199;
XX
DT   01-AUG-1991  (first entry)
XX
DE   Repeating unit.
XX
KW   Elastomeric copolymer; prosthesis; repeating unit.
XX
PN   US4589082-A.
XX
PD   20-MAY-1986.
XX
PF   19-SEP-1983;  83US-0533524.
XX
PR   19-SEP-1983;  83US-0533524.
XX
PA   (URRY/) URRY D W.
XX
PI   Urry DW;
XX
DR   WPI; 1986-149845/23.
XX
PT   Synthetic elastomeric copolymers - useful as prostheses for
PT   repair of ligaments, tendons and blood vessel walls
PS   Claim 22; page 10; 1ipp; English.
XX
CC   The repeating unit can be used in the prodn. of a synthetic elastic
CC   copolymer. The copolymer contains a beta turn and the repeating
CC   unit contains hydrophobic amino acid and glycine residues and a cross-
```

CC   linking component. The copolymer is useful in prosthetic systems, for
CC   repairing a natural elastic system. It is so functionalised so as to
CC   provide reactive gps. which can become covalently cross-linked by
CC   tissue enzymes to newly synthesised connective tissue protein.
XX
SQ   Sequence   5 AA;

Query Match   100.0%; Score 27; DB 7; Length 5;
Best Local Similarity   100.0%; Pred. No. 3.2e+05;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   1 vpgvg 5

RESULT 2
ID   R29149 standard; peptide; 5 AA.
XX
AC   R29149;
XX
DT   06-MAY-1993 (first entry)
XX
DE   Pentapeptide repeating unit of bioelastic polymer.
XX
KW   Superabsorbent; bioelastic; diaper; hygienic articles; wound;
KW   dressing; implant; inverse temperature transition; tissue; napkin;
KW   catbox liner; toilet paper; towellete; cleaning wipe; bandage;
KW   medical sponge; swab; printing ink; contact lens.
XX
OS   Synthetic.
XX
PN   WO9218079-A.
XX
PD   29-OCT-1992.
XX
PF   10-MAR-1992; 92WO-US01959.
XX
PR   19-APR-1991; 91US-0688185.
XX
PA   (BIOE-) BIOELASTICS RES LTD.
XX
PI   Urry DW;
XX
DR   WPI; 1992-381725/46.
XX
PT   Super-absorbent material incorporating polymer undergoing inverse
PT   temp. transition - esp. bio-elastic polypeptide(s) for
PT   controllably absorbing body fluids
XX
PS   Claim 27/28; Page 34; 42pp; English.
XX
CC   The invention relates to an appliance suitable for contacting body
CC   surfaces and for absorbing aqueous liquids including body fluids.
CC   The appliance includes a polymeric material which undergoes an
CC   inverse temperature transition, pref. such that it is in a
CC   contracted state at a higher temperature and in a swollen state when
CC   at a lower temperature. The appliance, e.g. a diaper, may be
CC   selected to be in a contracted state prior to use and while
CC   contacting the body surface, and to be in a swollen state after
CC   absorbing a body fluid and being at a new location having a lower
CC   temperature distant from the body-contacting location. Pref. the
CC   polymeric material is a bioelastic polymer containing elastomeric
CC   tetrapeptide, pentapeptide or nonapeptide repeat units, the polymer
CC   having a series of beta turns separated by dynamic bridging segments
CC   suspended between the beta turns. In particular, the polymer comprises a
CC   segment of formula poly[(VPGXG)x(VPGVG)y], in which x and y are mole
CC   fractions such that x+y = 1, and X is a hydrophobic amino acid residue;
CC   or of formula poly[(VPGXG)x(VPGVG)y(VPGZG)z], in which x+y+z = 1, X = as
CC   above, and 2 is an amino acid residue having a side chain capable of
CC   undergoing reversible protonation in an aqueous environment.

CC   A specific polymer is poly(VPGVG).
XX
SQ   Sequence   5 AA;

Query Match   100.0%; Score 27; DB 13; Length 5;
Best Local Similarity   100.0%; Pred. No. 3.2e+05;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   1 vpgvg 5

RESULT 3
ID   R29145 standard; peptide; 5 AA.
XX
AC   R29145;
XX
DT   23-MAR-1993 (first entry)
XX
DE   Bioelastic pentapeptide polymer fragment 2.
XX
KW   Polymer fragment; inverse temperature transition; exerted pressure;
KW   mechanical work; chemical change; hydrophobicity.
XX
OS   Synthetic.
XX
PN   WO9219183-A.
XX
PD   12-NOV-1992.
XX
PF   03-APR-1992; 92WO-US02691.
XX
PR   22-APR-1991; 91US-0688324.
XX
PA   (UABR-) UAB RES FOUND.
XX
PI   Urry DW;
XX
DR   WPI; 1992-398469/48.
XX
PT   Pressure expanding polymers giving reversible mechanical changes
PT   - has inverse temp. transition in liq. water range and pref.
PT   being bio-elastic polypeptide(s) contg. hydrophobic gps.
XX
PS   Claims 13 and 14; Page 36; 51pp; English.
XX
CC   The sequences given in R29144-46 are polymer fragments that are used
CC   in polymeric materials which have an inverse temp. transition in the
CC   range of liq. water and which expand against an exerted pressure.
CC   Fragments 1 and 2 are used such that mole fractions equal 1 in one
CC   polymeric material and fragments 1, 2 and 3 are used in a further
CC   polymeric material such that mole fractions equal 1. The materials
CC   can be used in a variety of applications to produce mechanical work
CC   and/or cause chemical changes in a sealed environment by variation of
CC   the pressure on the material. The degree of mechanical or chemical
CC   change can be controlled by selection of the number, hydrophobicity
CC   and size of the hydrophobic groups and the presence or absence of
CC   reactive functional groups in the polymer.
XX
SQ   Sequence   5 AA;

Query Match   100.0%; Score 27; DB 13; Length 5;
Best Local Similarity   100.0%; Pred. No. 3.2e+05;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   1 vpgvg 5

```
RESULT 4
R65228
ID   R65228 standard; peptide; 5 AA.
XX
AC   R65228;
XX
DT   12-OCT-1995 (first entry)
XX
DE   Fibrinogen/beta-casein transglutaminase cross-linkable copolymer.
XX
KW   Transglutaminase cross-linkable copolymer; fibrinogen; beta-casein;
KW   biocompatable; bioadhesive polymers; wound repair; prosthetics;
KW   bone and soft tissue matrices; controlled drug release carriers.
XX
OS   Synthetic.
XX
AC   WO9505396-A.
XX
DT   23-FEB-1995.
XX
PF   05-AUG-1994;   94WO-US08754.
XX
PR   13-AUG-1993;   93US-C106509.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Busby SJ, Labroo VM;
XX
DR   WPI; 1995-098722/13.
XX
PT   New peptide(s) based on fibrinogen and beta-casein -
PT   cross-linkable by trans:glutaminase, used for preparing
PT   biocompatible, bio:adhesive polymers
XX
PS   Claim 16; Page 48; 55pp; English.
XX
CC   A biocompatable, bioadhesive, transglutaminase cross-linkable
CC   copolymer comprises a first polypeptide monomer from 13-120 amino
CC   acids containing a segment of the formula S1-Y-S2 cross-linkable
CC   by a transglutaminase, where S1 = R65218, S2 = R65219 and Y = a
CC   spacer of 0-7 amino acids pref. R65220/21/22 or R65223, and
CC   a second polypeptide monomer selected from R65228-R65231 (the
CC   first monomer can also be R65227). The copolymer can be used to
CC   produce tissue adhesives, wound repair formulations, rigid
CC   prosthetics, matrices for the replacement of bone and soft tissue
CC   structures and carriers for controlled drug release compsns.
XX
SQ   Sequence   5 AA;

Query Match        100.0%;  Score 27;  DB 16;  Length 5;
Best Local Similarity  100.0%;  Pred. No. 3.2e+05;
Matches   5;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1  VPGVG 5
           |||||
Db      1  vpgvg 5

RESULT 5
R80250
ID   R80250 standard; peptide; 5 AA.
XX
AC   R80250;
XX
DT   17-APR-1996 (first entry)
XX
DE   Elastin-like repetative unit.
XX
KW   fibroin; elastin; repeat sequence; suture; thread; pin; gel; silk;
KW   polymer.
XX
```

```
OS   Synthetic.
XX
PN   WO9524478-A1.
XX
PD   14-SEP-1995.
XX
PF   10-MAR-1995;   95WO-US02772.
XX
PR   11-MAR-1994;   94US-0212237.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J;
XX
DR   WPI; 1995-328270/42.
XX
PT   Protein polymer comprising alternating blocks of fibroin and elastin
PT   units - used to form a device e.g. a suture to keep separated
PT   viable tissue together
XX
PS   Claim 1; Page 41; 46pp; English.
XX
CC   This sequence represents a repetative unit based on elastin. This
CC   sequence is used in combination with a repetative unit based on fibroin
CC   (silk) (see R80249). Polymers were constructed that consisted of
CC   alternating blocks of these two sequences (see R80251-R80257). The
CC   repeating units of the polymers contained 2-12 units of the fibroin
CC   sequence, and 2-32 units of this sequence. The polymers are used to form
CC   a device (such as a suture, pin, thread, gel or film) to keep separated
CC   viable tissue together. By varying the the ratio of the two repetative
CC   units, and by altering the lengths of the blocks of each of them, the
CC   tensile properties of the polymer can altered moderatly. By reducing the
CC   number of repeating units of this sequence, or by increasing the number
CC   of units of the elastin like repeat, a faster rate of resorption can be
CC   achieved.
XX
SQ   Sequence   5 AA;

Query Match        100.0%;  Score 27;  DB 16;  Length 5;
Best Local Similarity  100.0%;  Pred. No. 3.2e+05;
Matches   5;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1  VPGVG 5
           |||||
Db      1  vpgvg 5

RESULT 6
R80307
ID   R80307 standard; peptide; 5 AA.
XX
AC   R80307;
XX
DT   17-APR-1996 (first entry)
XX
DE   Elastin cross-linking substrate consensus sequence.
XX
KW   Pendent group; repeating unit; enzyme recognition site; sealant; elastin;
KW   enzymatic cross-linking; biocompatible material; structural integrity;
KW   medical adhesive; wound closure; tissue repair.
XX
OS   Synthetic.
XX
PN   WO9523611-A1.
XX
PD   08-SEP-1995.
XX
PF   03-MAR-1995;   95WO-US02728.
XX
PR   03-MAR-1994;   94US-0205518.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
```

```
XX
PI   Cappello J;
XX
DR   WPI; 1995-320413/41.
XX
PT   Protein polymers comprising repeating units and sequences - capable
PT   of enzyme-catalysed covalent bond formation useful as a
PT   biocompatible material for wound closure and tissue repair.
XX
PS   Disclosure; Page 12; 138pp; English.
XX
CC   The sequence of the cross-linking reactive motif from elastin. The motif
CC   can be used in a novel polymer comprising two spaced enzyme recognition
CC   site and may contain repetitive units of 3-8 amino acids with at least
CC   two pendent groups. The polymers contg. the multimeric repeat sequence
CC   can be used as substrates for enzymatic cross-linking. The polymers can
CC   be used in biological systems where in situ formation of a biocompatible
CC   material with structural integrity is required e.g. as medical adhesives
CC   and sealants or for wound closure or tissue repair.
XX
SQ   Sequence  5 AA;

Query Match      100.0%;  Score 27;  DB 16;  Length 5;
Best Local Similarity  100.0%;  Pred. No. 3.2e-05;
Matches   5;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1 VPGVG 5
       |||||
Db   1 vpgvg 5


RESULT   7
W22714
ID   W22714 standard; peptide; 5 AA.
XX
AC   W22714;
XX
DT   26-FEB-1998  (first entry)
XX
DE   Bioelastomeric repeating unit 1 responsive to electrical energy.
XX
KW   Bioelastomeric repeating unit; polymer; inverse temperature transition;
KW   electrical energy change; polarity; hydrophobicity; mechanical work;
KW   surgical suture; heat shrinkable membrane; desalination.
OS   Synthetic.
XX
PN   WO9723729-A1.
XX
PD   03-JUL-1997.
XX
PF   07-JUN-1996;  96WO-US09775.
PR   07-JUN-1995;  95US-0487594.
PA   (URRY/) URRY D W.
PI   Urry DW;
XX
DR   WPI; 1997-363350/33.
XX
PT   Bioelastic polymer responsive to electrical energy - comprising
PT   beta turn and residue(s) with side chain that changes polarity or
PT   hydrophobicity in response to electrical energy change, useful for
PT   mechanical work or light stimulated contraction
XX
PS   Claim 13; Page 52; 60pp; English.
XX
CC   W22711-16 represent bioelastomeric repeating units that are used to
CC   create a polymer responsive to electrical energy. This bioelastic
CC   polymer comprises a bioelastomeric polypeptide having an inverse
CC   temperature transition, and a repeating bioelastomeric unit containing at
CC   least 1 beta-turn and residues with a side chain. The side chain of these
CC   residues responds to an electrical energy change by altering its polarity
CC   or hydrophobicity. The side chain is present in sufficient amounts to
CC   provide a shift in the inverse temperature transition of the polymer upon
CC   the change in exposure to electrical energy. The polymer may also contain
CC   a second amino acid with a side chain capable of undergoing a change in
CC   an aqueous environment. The polymer (W34881-83) consists of the formulas
CC   described in W22716 and W22711 and below:
CC   poly[fx(VPGVG),fv(VPGVG)] where
CC   fx and fv are mole fractions with fx + fv = 1
CC   X represents an amino acid residue having an electrically responsive
CC   side chain.
CC   The expansion or contraction of the polymer upon exposure to a change
CC   in electrical energy can be used to produce mechanical work when the
CC   polymer is constrained. It can also cause turbidity and chemical changes
CC   in a closed environment. The polymer can be used in surgical sutures,
CC   especially for microsurgery, heat shrinkable membranes, controlled
CC   herbicide or pesticide release or for desalination.
XX
SQ   Sequence  5 AA;

Query Match      100.0%;  Score 27;  DB 18;  Length 5;
Best Local Similarity  100.0%;  Pred. No. 3.2e+05;
Matches   5;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1 VPGVG 5
       |||||
Db   1 vpgvg 5


RESULT   8
W26332
ID   W26332 standard; Peptide; 5 AA.
XX
AC   W26332;
XX
DT   19-NOV-1997  (first entry)
XX
DE   Elastin repeat unit.
XX
KW   Elastin; protein polymer.
OS   Synthetic.
XX
PN   US5641648-A.
XX
PD   24-JUN-1997.
XX
PF   04-NOV-1986;  86US-0927258.
PR   29-DEC-1993;  93US-0175155.
PR   04-NOV-1986;  86US-0927258.
PR   29-OCT-1987;  87US-0114618.
PR   09-NOV-1988;  88US-0269429.
PR   06-NOV-1990;  90US-0609716.
PR   22-APR-1993;  93US-0053049.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J, Ferrari FA, Richardson C;
XX
DR   WPI; 1997-340943/31.
XX
PT   Preparation of synthetic DNA encoding a protein comprising short
PT   repeats - by synthesising oligomers, annealing and oligomerising
PT   these, particularly to produce proteins that mimic silk, collagen
PT   etc
XX
PS   Claim 8; Column 11; 90pp; English.
XX
CC   This peptide represents a repeat unit found in elastin. Methods
CC   are claimed for preparing protein polymers that contain repeating
```

CC units based on those found in elastin, fibroin, keratin, etc. (see
CC also W26329, W26334 and W26336). The method involves: synthesising
CC pairs of single-stranded oligomers, each member of the pair
CC overlapping except at the protruding ends; hybridising each pair to
CC double-stranded segments; combining the double-stranded segments, or
CC their cloned copies, in a cloning vector to form a monomer; excising
CC the monomer from the vector; and oligomerising the monomer to
CC produce a multimer. Such polymers comprise fibrous or structural
CC proteins, including crystalline, elastomeric, tough and bony
CC materials such as those that mimic (and can substitute for) silk,
CC elastin, collagen, keratin etc. The properties of the protein can
CC be controlled by varying the type of units in the monomer, the
CC number of units per multimer, the spacing between them and the
CC number of multimer repeats.
XX
SQ Sequence 5 AA;

Query Match   100.0%; Score 27; DB 18; Length 5;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   1 vpgvg 5

RESULT 9
ID   W18261 standard; peptide; 5 AA.
XX
AC   W18261;
XX
DT   29-AUG-1997 (first entry)
XX
DE   Transglutaminase cross-linkable polypeptide elastomeric peptide.
XX
KW   Elastomeric; homopolymer; copolymer; tissue sealant; skin graft;
KW   burn.
XX
OS   Synthetic.
XX
PN   WO9640780-A1.
XX
PD   19-DEC-1996.
XX
PF   31-MAY-1996; 96WO-US08269.
XX
PR   07-JUN-1995; 95US-0483236.
XX
PA   (ZYMO) ZYMOGENETICS INC.
XX
PI   Busby SJ, Labroo VM;
XX
DR   WPI; 1997-052237/05.
XX
PT   Trans:glutaminase cross-linkable peptide(s) - used in the mfr. of
PT   biocompatible, bio:adhesive tissue sealant and wound healing
PT   preparations.
XX
PS   Claim 6; Page 48; 55pp; English.
XX
CC   Novel polypeptides optionally having one or both of the amino-terminus
CC   and carboxy-terminus flanked by an elastomeric peptide, which is cross-
CC   linkable by a transglutaminase comprises a segment of formula:
CC   S1-Y-S2. The present sequence represents a specifically claimed
CC   example of an elastomeric peptide. The homo- and copolymers produced
CC   are useful in tissue sealant and wound healing formulations. Tissue
CC   sealants are useful in skin grafting for burn victims and for sealing
CC   surgical and other wounds.
XX
SQ Sequence 5 AA;

Query Match   100.0%; Score 27; DB 18; Length 5;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   1 vpgvg 5

RESULT 10
ID   W12301 standard; peptide; 5 AA.
XX
AC   W12301;
XX
DT   22-APR-1997 (first entry)
XX
DE   Monomeric unit for elastic protein-base polymer for chewing gum.
XX
KW   Chewing gum; water; insoluble; plasticiser; elastomer; flavour; polymer;
KW   sweetener; texturiser; crosslinkage; gamma-irradiation; drug; nutrient;
KW   environmentally friendly; drug; anti-inflammatory agent; vitamin.
XX
OS   Synthetic.
XX
PN   US5580590-A.
XX
PD   03-DEC-1996.
XX
PF   27-DEC-1993; 93US-0174185.
XX
PR   27-DEC-1993; 93US-0174185.
XX
PA   (WRIL) WRIGLEY JR CO WM.
XX
PI   Hartman SE;
XX
DR   WPI; 1997-033531/03.
XX
PT   Environmentally friendly chewing gum - contg. water insoluble
PT   elastic polypeptide having a penta:peptide repeat
XX
PS   Claim 1; Column 7; 6pp; English.
XX
CC   A novel chewing gum composition comprises a water insoluble polymer
CC   comprising this pentapeptide as a base. The insoluble portion can
CC   comprise 1-99% of the gum, together with a plasticiser, an elastomer,
CC   a flavour, a sweetener and a texturiser. The peptide is preferably
CC   crosslinked by gamma-irradiation to form the polymer. The new chewing
CC   gum is environmentally friendly as it can be swallowed after chewing or
CC   easily removed from surfaces. The peptide can also have drugs or other
CC   nutrients e.g. anti-inflammatory agents, vitamins, etc., attached to the
CC   peptide, for therapeutic purposes.
XX
SQ Sequence 5 AA;

Query Match   100.0%; Score 27; DB 18; Length 5;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   1 vpgvg 5

RESULT 11
ID   W49701 standard; Peptide; 5 AA.
XX
AC   W49701;
XX

```
DT   12-OCT-1998  (first entry)
XX
DE   Elastin repeat motif.
XX
KW   Protein polymer; cross-linking; elastin; adhesive; sealant;
KW   wound healing.
XX
OS   Synthetic.
XX
PN   US5773577-A.
XX
PD   30-JUN-1998.
XX
PF   03-MAR-1994;   94US-0205518.
XX
PR   02-MAR-1995;   95US-0397633.
PR   03-MAR-1994;   94US-0205518.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J;
XX
DR   WPI; 1998-387091/33.
XX
PT   New recombinant protein polymers - containing naturally occurring
PT   repetitive units for crosslinking by enzymes, useful as medical
PT   adhesives and sealants, depots and matrices
XX
PS   Disclosure; Column 7; 70pp; English.
XX
CC   This is a repeat motif of elastin. A claimed recombinant protein
CC   polymer of 15-250 kDa comprises a repetitive amino acid backbone of
CC   repetitive units having a collagen, fibroin, elastin or keratin
CC   motif and at least 2 enzyme recognition sequences comprising a
CC   glutamine capable of enzyme-catalysed isopeptide formation,
CC   separated by an intervening sequence of at least 25 amino acids.
CC   Such recombinant protein polymers are capable of covalent
CC   crosslinking by enzymatic reaction to form products which set
CC   quickly and have good adhesive properties and high strength. The
CC   compositions can be used as medical adhesives and sealants, in the
CC   closure of wounds and repair of damaged tissues, prosthesis
CC   coatings, drug depots, and matrices for the transplantation of
CC   cells. They can also be used in assays for analytes.
XX
SQ   Sequence   5 AA;

Query Match     100.0%;  Score 27;  DB 19;  Length 5;
Best Local Similarity  100.0%;  Pred. No. 3.2e+05;
Matches    5;  Conservative    0;  Mismatches   0;  Indels    0;  Gaps   0;

Qy          1 VPGVG 5
              |||||
Db          1 VPgvg 5


RESULT  12
W53506
ID   W53506 standard; Protein; 5 AA.
XX
AC   W53506;
XX
DT   10-AUG-1998  (first entry)
XX
DE   Peptide reapeat unit 4.
XX
KW   Peptide repeat unit; DNA repeat unit; high molecular weight polymer;
KW   synthetic silk; silk worm; elastin.
XX
OS   Synthetic.
OS   Bombyx mori.
XX
PN   WO9810063-A1.
XX
```

```
XX
PD   12-MAR-1998.
XX
PF   23-SEP-1996;   96WO-US15306.
XX
PR   03-SEP-1996;   96US-0707237.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J,  Crissman JW,  Dorman MA,  Ferrari FA;
XX
DR   WPI; 1998-193613/17.
XX
PT   Preparation of synthetic repetitive DNA - useful for construction of
PT   large protein polymers having repeating units, used in structural
PT   material, e.g. synthetic silk
XX
PS   Claim 11; Page 22; 127pp; English.
XX
CC   This is the amino acid sequence of the peptide repeat unit, which
CC   is generally found in elastin, and used in the method of the invention,
CC   which involves the preparation of synthetic DNA sequence having
CC   repeating units from about 3-15 codons and encoding a protein of at
CC   least about 30 kDa. The method is useful for the production of high
CC   molecular weight polymers (e.g. synthetic silk), either nucleic acids
CC   or peptides that are the expression products of the nucleic acids and
CC   particularly high molecular weight peptides containing repeating units
CC   which are useful as structural materials.
XX
SQ   Sequence   5 AA;

Query Match     100.0%;  Score 27;  DB 19;  Length 5;
Best Local Similarity  100.0%;  Pred. No. 3.2e+05;
Matches    5;  Conservative    0;  Mismatches   0;  Indels    0;  Gaps   0;

Qy          1 VPGVG 5
              |||||
Db          1 vpgvg 5


RESULT  13
W47348
ID   W47348 standard; peptide; 5 AA.
XX
AC   W47348;
XX
DT   01-JUN-1998  (first entry)
XX
DE   Elastin repeat unit.
XX
KW   Binding inhibitor; low-density lipoprotein; LDL; vascular wall;
KW   vascular injury; elastin; collagen; prevention; treatment;
KW   vascular disease; atherosclerosis; repeat unit.
XX
OS   Homo sapiens.
XX
PN   US5726153-A.
XX
PD   10-MAR-1998.
XX
PF   05-JUN-1995;   95US-0468543.
XX
PR   06-JUN-1995;   95US-0469692.
PR   02-MAY-1988;   88US-0189130.
PR   03-MAY-1990;   90US-0518142.
PR   03-MAY-1990;   90US-0518215.
PR   02-MAY-1991;   91US-0694929.
PR   16-APR-1993;   93US-0048569.
PR   24-FEB-1994;   94US-0201057.
PR   28-FEB-1995;   95US-0398046.
PR   06-JUN-1995;   95US-0468543.
XX
```

PA (NEWE-) NEW ENGLAND DEACONESS HOSPITAL.
XX
PI Findeis MA, Fischman A, Lees AM, Lees RS, Shih I;
XX
DR WPI: 1998-192802/17.
XX
PT Inhibiting binding of low-density lipoprotein to vascular walls - by administering peptide with affinity for vascular injury sites
XX
PS Disclosure: Column 14; 31pp; English.
XX
CC A novel method for inhibiting the binding of a low-density lipoprotein (LDL) to vascular walls in vivo, comprises administering a synthetic water soluble peptide containing an amphiphilic domain and having affinity for sites of vascular injury, e.g. derived from the present peptide. The peptide inhibits the binding of LDL to vascular wall components, e.g. elastin and collagen, and so can be used to prevent or treat vascular diseases, e.g. atherosclerosis.
XX
SQ Sequence 5 AA;

Query Match 100.0%; Score 27; DB 19; Length 5;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPGVG 5
      |||||
Db 1 vpgvg 5

RESULT 14
Y55877
ID Y55877 standard; peptide; 5 AA.
XX
AC Y55877;
XX
DT 01-FEB-2000 (first entry)
XX
DE Apolipoprotein fragment peptide #27 for vascular disease imaging.
XX
KW Diagnosis; water soluble; amphiphilic domain; affinity; vascular injury; detection; imaging; disease; atherosclerosis; apolipoprotein.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN US5972890-A.
XX
PD 26-OCT-1999.
XX
PF 28-FEB-1995; 95US-0398046.
XX
PR 02-MAY-1991; 91US-0694929.
PR 16-APR-1993; 93US-0048569.
PR 24-FEB-1994; 94US-0201057.
PR 24-FEB-1994; 88US-0189130.
PR 03-MAY-1990; 90US-0518142.
PR 03-MAY-1990; 90US-0518215.
XX
PA (NEWE-) NEW ENGLAND DEACONESS HOSPITAL.
XX
PI Lees AM, Fischman A, Shih I, Findeis MA, Lees RS;
XX
DR WPI: 1999-632641/54.
XX
PT New diagnostic synthetic peptides which have affinity for and accumulate at a site of vascular injury useful for detection and imaging of vascular disease such as atherosclerosis -
XX
PS Claim 37; Column 34; 30pp; English.
XX

CC The peptides Y55851-Y55889 represent examples of diagnostic, synthetic peptides which carry a detectable label, contain 30 or fewer amino acids, are water soluble, contain an amphiphilic domain and have affinity for, and propensity to accumulate at, a site of vascular injury. They are preferably derived from the amino acid sequence of apolipoprotein. The peptides can be used for the detection or imaging of a vascular injury or disease, e.g. atherosclerosis.
XX
SQ Sequence 5 AA;

Query Match 100.0%; Score 27; DB 20; Length 5;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPGVG 5
      |||||
Db 1 vpgvg 5

RESULT 15
Y31686
ID Y31686 standard; Peptide; 5 AA.
XX
AC Y31686;
XX
DT 22-NOV-1999 (first entry)
XX
DE Pentapeptide used in novel elastomer polymers.
XX
KW Elastomer; bioelastomer; polymer; tissue augmentation; tissue restoration; tissue reconstruction; tissue repair; implant.
XX
OS Synthetic.
XX
PN WO9943271-A1.
XX
PD 02-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04440.
XX
PR 29-MAY-1998; 98US-0087155.
PR 27-FEB-1998; 98US-0076297.
XX
PA (BIOE-) BIOELASTICS RES LTD.
XX
PI Glazer PA, Parker TM, Urry DW;
XX
DR WPI: 1999-540487/45.
XX
PT Augmentation or restoration of mammalian tissue by injecting solution of peptide polymer, used for soft or hard tissue reconstruction, especially of intervertebral disks
XX
PS Claim 9; Page 75; 133pp; English.
XX
CC The present sequence represents an example of a pentapeptide that is used in novel bioelastic polymers. The invention provides a method of tissue augmentation by injecting a polymer comprising repeating peptide monomeric units selected from nonapeptide, pentapeptide and tetrapeptide monomeric units, where the monomeric units form a series of beta-turns separated by dynamic bridging segments. The polymer has an inverse temperature transition value that is less than the tissue temp. and is injected in water solution at coacervate concentration. The polymer can be injected at periurethral or subdermal sites (for treatment of urinary incontinence or for cosmetic purposes), or into hard or soft tissue, e.g. for repair of traumatic injury. A specific application is restoration of intervertebral discs.
XX
SQ Sequence 5 AA;

Query Match   100.0%; Score 27; DB 20; Length 5;
Best Local Similarity   100.0%; Pred. No. 3.2e+05; Indels   0; Gaps   0;
Matches   5; Conservative   0; Mismatches   0;

Qy   1 VPGVG 5
         |||||
Db   1 vpgvg 5

Search completed: April 24, 2001, 16:38:19
Job time: 419 sec

```
            GenCore version 4.5
       Copyright (c) 1993 - 2000 Compugen Ltd.


OM protein - protein search, using sw model

Run on:          April 24, 2001, 16:36:24 ; Search time 62.39 seconds
                                            (without alignments)
                                            1.540 Million cell updates/sec

Title:           US-09-340-736-6
Perfect score:   27
Sequence:        1 VPGVG 5

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Database:        185757 segs, 1920857 residues

Total number of hits satisfying chosen parameters:    185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
          1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
          2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
          3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
          4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
          5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
          6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	27	100.0	5	1	US-08-106-509-5	Sequence 5, Appli
2	27	100.0	5	1	US-07-603-716-2	Sequence 2, Appli
3	27	100.0	5	1	US-08-174-185-1	Sequence 1, Appli
4	27	100.0	5	1	US-08-212-237-2	Sequence 2, Appli
5	27	100.0	5	1	US-08-175-155-4	Sequence 4, Appli
6	27	100.0	5	1	US-08-175-155-9	Sequence 9, Appli
7	27	100.0	5	1	US-08-468-543-15	Sequence 15, Appli
8	27	100.0	5	1	US-08-477-509B-24	Sequence 24, Appli
9	27	100.0	5	1	US-08-397-633A-1	Sequence 1, Appli
10	27	100.0	5	1	US-08-397-633A-8	Sequence 8, Appli
11	27	100.0	5	2	US-08-707-237A-10	Sequence 10, Appli
12	27	100.0	5	2	US-08-483-236-15	Sequence 15, Appli
13	27	100.0	5	2	US-08-469-692-15	Sequence 15, Appli
14	27	100.0	5	2	US-08-911-364-6	Sequence 6, Appli
15	27	100.0	5	2	US-08-735-692-20	Sequence 20, Appli
16	27	100.0	5	2	US-08-398-046-15	Sequence 15, Appli
17	27	100.0	5	3	US-08-542-051-1	Sequence 1, Appli
18	27	100.0	5	3	US-08-482-085B-24	Sequence 24, Appli
19	27	100.0	5	3	US-08-482-085B-43	Sequence 43, Appli
20	27	100.0	5	4	US-08-963-168C-27	Sequence 27, Appli
21	27	100.0	5	4	US-08-475-411A-2	Sequence 2, Appli
22	27	100.0	5	4	US-08-478-029A-2	Sequence 2, Appli
23	27	100.0	5	5	PCR-US95-02772-2	Sequence 2, Appli
24	27	100.0	5	6	5250516-1	Patent No. 5250516
25	27	100.0	5	6	5250516-17	Sequence 17, Appli
26	27	100.0	6	1	US-08-468-543-12	Sequence 12, Appli
27	27	100.0	6	2	US-08-469-692-12	Sequence 12, Appli
28	27	100.0	6	2	US-08-398-046-12	Sequence 12, Appli
29	27	100.0	11	6	5250516-6	Patent No. 5250516
30	27	100.0	16	2	US-08-468-543-17	Sequence 17, Appli
31	27	100.0	16	2	US-08-469-692-17	Sequence 17, Appli
32	27	100.0	16	2	US-08-398-046-17	Sequence 17, Appli
33	27	100.0	20	1	US-08-127-351-35	Sequence 35, Appli
34	27	100.0	20	1	US-08-480-367B-35	Sequence 35, Appli
35	27	100.0	20	1	US-08-487-221A-35	Sequence 35, Appli
36	27	100.0	20	1	US-08-480-370-35	Sequence 35, Appli
37	27	100.0	20	1	US-08-299-536-14	Sequence 14, Appli
38	27	100.0	20	1	US-08-279-155-16	Sequence 16, Appli
39	27	100.0	20	1	US-08-464-455-14	Sequence 14, Appli
40	27	100.0	20	1	US-08-486-135-2	Sequence 2, Appli
41	27	100.0	20	1	US-08-703-988A-16	Sequence 16, Appli
42	27	100.0	20	1	US-08-470-152-2	Sequence 2, Appli
43	27	100.0	20	1	US-08-463-052-14	Sequence 14, Appli
44	27	100.0	20	2	US-08-480-551-14	Sequence 14, Appli
45	27	100.0	20	2	US-08-612-842-16	Sequence 16, Appli

ALIGNMENTS

```
RESULT 1
US-08-106-509-5
; Sequence 5, Application US/08106509
; Patent No. 5428014
; GENERAL INFORMATION:
;   APPLICANT: Labroo, Virender
;   APPLICANT: Busby, Sharon J.
;   TITLE OF INVENTION: TRANSGLUTAMINASE CROSS-LINKABLE
;   TITLE OF INVENTION: POLYPEPTIDES AND METHODS RELATING THERETO
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: ZymoGenetics, Inc.
;     STREET: 4225 Roosevelt Way, N.E.
;     CITY: Seattle
;     STATE: WA
;     COUNTRY: USA
;     ZIP: 98105
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/106,509
;     FILING DATE: 19930813
;     CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Parker, Gary E
;     REGISTRATION NUMBER: 31-684
;     REFERENCE/DOCKET NUMBER: 93-09
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 206-547-8080 ext 322
;     TELEFAX: 206-548-2329
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 5 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   FRAGMENT TYPE: internal
US-08-106-509-5

  Query Match             100.0%;  Score 27;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+05;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VPGVG 5
              |||||
Db          1 VPGVG 5
```

RESULT 2
US-07-609-716-2
; Sequence 2, Application US/07609716
; Patent No. 551581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-609-716-2

Query Match   100.0%; Score 27; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPGVG 5
        |||||
Db 1 VPGVG 5

RESULT 3
US-08-174-185-1
; Sequence 1, Application US/08174185
; Patent No. 5580590
; GENERAL INFORMATION:
; APPLICANT: Hartman, Scott E
; TITLE OF INVENTION: Environmentally Friendly
; TITLE OF INVENTION: Chewing Gum Compositions Containing Elastic Protein-Based
; TITLE OF INVENTION: Polymer
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hill, Steadman & Simpson
; STREET: 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/174,185
; FILING DATE: 27-DEC-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acid residues
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
; MOLECULE TYPE: peptide
US-08-174-185-1

Query Match   100.0%; Score 27; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPGVG 5
        |||||
Db 1 VPGVG 5

RESULT 4
US-08-212-237-2
; Sequence 2, Application US/08212237
; Patent No. 5605019
; GENERAL INFORMATION:
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Synthetic Proteins As Implantables
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,237
; FILING DATE: 11-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-58847/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-212-237-2

Query Match   100.0%; Score 27; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPGVG 5
        |||||

Db   1 VPGVG 5

RESULT 5
US-08-175-155-4
; Sequence 4, Application US/08175155
; Patent No. 5641648
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; APPLICANT: Crissman, John W.
; APPLICANT: Dorman, Mary A.
; TITLE OF INVENTION: Methods for Preparing Synthetic
; TITLE OF INVENTION: Repetitive DNA
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/175,155
; FILING DATE: 29-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-5/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-175-155-4

Query Match   100.0%; Score 27; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   5; Conservative   0; Mismatches   0; Indels 0; Gaps 0;

Qy   1 VPGVG 5
        |||||
Db   1 VPGVG 5

RESULT 6
US-08-175-155-9
; Sequence 9, Application US/08175155
; Patent No. 5641648
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; APPLICANT: Crissman, John W.
; APPLICANT: Dorman, Mary A.
; TITLE OF INVENTION: Methods for Preparing Synthetic
; TITLE OF INVENTION: Repetitive DNA
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRSS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/175,155
; FILING DATE: 29-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-5/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-175-155-9

Query Match   100.0%; Score 27; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   5; Conservative   0; Mismatches   0; Indels 0; Gaps 0;

Qy   1 VPGVG 5
        |||||
Db   1 VPGVG 5

RESULT 7
US-08-468-543-15
; Sequence 15, Application US/08468543
; Patent No. 5726153
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S. et al.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR ARTERIAL IMAGING
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,543
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/398,046
; FILING DATE: 02-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/201,057
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,569
; FILING DATE: 16-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694,929

FILING DATE: 02-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,215
FILING DATE: 03-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/518,142
FILING DATE: 03-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/189,130
FILING DATE: 02-MAY-1988
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 04547/002003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 5
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-468-543-15

Query Match   100.0%; Score 27; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPGVG 5
       |||||
Db 1 VPGVG 5

RESULT 8
US-08-477-509B-24
; Sequence 24, Application US/08477509B
; Patent No. 5770597
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A
APPLICANT: Cappello, Joseph
APPLICANT: Crissman, John W
APPLICANT: Dorman, Mary A
TITLE OF INVENTION: No. 5770697el Peptides Comprising Repetitive
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
NUMBER OF SEQUENCES: 112
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,509B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,155
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,049
FILING DATE: 22-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-7/RFT/AGTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-477-509B-24

Query Match   100.0%; Score 27; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPGVG 5
       |||||
Db 1 VPGVG 5

RESULT 9
US-08-397-633A-1
; Sequence 1, Application US/08397633A
; Patent No. 5773377
GENERAL INFORMATION:
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE
TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING
NUMBER OF SEQUENCES: 105
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/397,633A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-397-633A-1

Query Match   100.0%; Score 27; DB 1; Length 5;

Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1 VPGVG 5
       |||||
Db   1 VPGVG 5
```

RESULT 10
US-08-397-633A-8
; Sequence 8, Application US/08397633A
; Patent No. 5773577
; GENERAL INFORMATION:
;   APPLICANT: Cappello, Joseph
;   TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE
;   TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING
;   NUMBER OF SEQUENCES: 105
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;   STREET: 4 Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/397,633A
;   FILING DATE:
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Rowland, Bertram I
;   REGISTRATION NUMBER: 20,015
;   REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROF-011-1
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 781-1989
;   TELEFAX: (415) 398-3249
;   TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 5 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-397-633A-8

Query Match 100.0%; Score 27; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1 VPGVG 5
       |||||
Db   1 VPGVG 5
```

RESULT 11
US-08-707-237A-10
; Sequence 10, Application US/08707237A
; Patent No. 5830713
; GENERAL INFORMATION:
;   APPLICANT: Ferrari, Franco A.
;   APPLICANT: Capello, Joseph
;   APPLICANT: Crissman, John W.
;   APPLICANT: Dorman, Mary A.
;   TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC
;   TITLE OF INVENTION: REPETITIVE DNA
;   NUMBER OF SEQUENCES: 108
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: Four Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: United States
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/707,237A
;   FILING DATE: 03-SEP-1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/175,155
;   FILING DATE: 29-DEC-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/053,049
;   FILING DATE: 22-APR-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/609,716
;   FILING DATE: 06-NOV-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/259,429
;   FILING DATE: 09-NOV-1988
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/114,618
;   FILING DATE: 29-OCT-1987
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 06/927,258
;   FILING DATE: 04-NOV-1986
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Trecartin, Richard F.
;   REGISTRATION NUMBER: 31,801
;   REFERENCE/DOCKET NUMBER: A-55186-10/WHD
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 781-1989
;   TELEFAX: (415) 398-3249
;   TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 5 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: unknown
;   TOPOLOGY: unknown
;   MOLECULE TYPE: protein
US-08-707-237A-10

Query Match 100.0%; Score 27; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1 VPGVG 5
       |||||
Db   1 VPGVG 5
```

RESULT 12
US-08-483-236-15
; Sequence 15, Application US/08483236
; Patent No. 5593385
; GENERAL INFORMATION:
;   APPLICANT: Labrou, Virender
;   APPLICANT: Busby, Sharron
;   TITLE OF INVENTION: Transglutaminase Cross-Linkable
;   TITLE OF INVENTION: Polypeptides and Methods Relating Thereto
;   NUMBER OF SEQUENCES: 18
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: ZymoGenetics, Inc.
;   STREET: 1201 Eastlake Avenue East

; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,236
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 93-09c1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6678
; TELEFAX: 206-442-6673
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-483-236-15

Query Match   100.0%; Score 27; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
        |||||
Db   1 VPGVG 5

RESULT 13
US-08-469-692-15
; Sequence 15, Application US/08469692
; Patent No. 5955055
; GENERAL INFORMATION:
; APPLICANT: Lees, Robert S. et al.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR ARTERIAL IMAGING
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,692
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/398,046
; FILING DATE: 02-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/201,057
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,569
; FILING DATE: 16-APR-1993
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/694,929
; FILING DATE: 02-MAY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/518,215
; FILING DATE: 03-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/518,142
; FILING DATE: 03-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/189,130
; FILING DATE: 02-MAY-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04547/002002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-469-692-15

Query Match   100.0%; Score 27; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
        |||||
Db   1 VPGVG 5

RESULT 14
US-08-911-364-6
; Sequence 6, Application US/08911364
; Patent No. 5969106
; GENERAL INFORMATION:
; APPLICANT: ROTHSTEIN, Aser
; APPLICANT: KEELY, Fred W.
; APPLICANT: ROTHSTEIN, Steven J.
; TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
; TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,364
; FILING DATE: 07-AUG-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,552
; FILING DATE: 07-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 041082/0104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300

;   TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 5 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-911-364-6

```
Query Match      100.0%;  Score 27;  DB 2;  Length 5;
Best Local Similarity  100.0%;  Pred. No. 1.4e+05;
Matches    5;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 VPGVG 5
            |||||
          1 VPGVG 5
```

RESULT 15
US-08-735-692-20
; Sequence 20, Application US/08735692B
; Patent No. 5972406
; GENERAL INFORMATION:
;   APPLICANT: Urry,Dan W.
;   APPLICANT: Sherry, Peter R.
;   APPLICANT: Prasad, Kari
;   TITLE OF INVENTION: Bioelastomers Suitable as Food Product Additives
;   FILE REFERENCE: BERL-017/01US
;   CURRENT APPLICATION NUMBER: US/08/735,692B
;   CURRENT FILING DATE: 1995-10-16
;   EARLIER APPLICATION NUMBER: 08/423,517
;   EARLIER FILING DATE: 1995-04-14
;   NUMBER OF SEQ ID NOS: 31
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:synthetic
US-08-735-692-20

```
Query Match      100.0%;  Score 27;  DB 2;  Length 5;
Best Local Similarity  100.0%;  Pred. No. 1.4e+05;
Matches    5;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 VPGVG 5
            |||||
Db        1 VPGVG 5
```

Search completed: April 24, 2001, 16:36:24
Job time: 304 sec

```
                GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        April 24, 2001, 16:42:01 ; Search time 74.56 Seconds
                                          (without alignments)
                                          4.609 Million cell updates/sec

Title:         US-09-340-736-6
Perfect score: 27
Sequence:      1 VPGVG 5

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters:  198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     PIR_67.*
               1: pir1.*
               2: pir2.*
               3: pir3.*
               4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	27	100.0	38	2	T46593	phytoene dehydroge
2	27	100.0	76	2	I45885	elastin - bovine (
3	27	100.0	93	2	C40595	hypothetical prote
4	27	100.0	120	2	A75586	hydrogenase expres
5	27	100.0	127	2	A72712	hypothetical prote
6	27	100.0	139	2	C70680	hypothetical prote
7	27	100.0	170	2	B83564	hypothetical prote
8	27	100.0	181	2	T05925	hypothetical prote
9	27	100.0	195	2	G64359	adenylate kinase (
10	27	100.0	199	2	E59534	DNA polymerase, ba
11	27	100.0	204	2	E72695	adenylate kinase (
12	27	100.0	206	2	D70544	probable amidotran
13	27	100.0	206	2	T45249	probable amidotran
14	27	100.0	213	2	T36699	probable two-compo
15	27	100.0	214	2	T23593	hypothetical prote
16	27	100.0	222	2	JQ0640	glutamine amidotra
17	27	100.0	231	2	D72532	hypothetical prote
18	27	100.0	246	2	T37169	hypothetical prote
19	27	100.0	255	2	A70778	hypothetical prote
20	27	100.0	271	2	S76871	hypothetical prote
21	27	100.0	274	2	G70898	probable pyrr prot
22	27	100.0	274	2	B49930	orotidine-5'-phosp
23	27	100.0	281	2	G70734	hypothetical prote
24	27	100.0	283	2	B38604	poly(3-hydroxyalka
25	27	100.0	285	2	S29306	poly(3-hydroxyalka
26	27	100.0	293	2	B75441	conserved hypothet
27	27	100.0	294	2	E59968	conserved hypothet
28	27	100.0	303	1	S75782	methanol dehydroge
29	27	100.0	305	2	E75308	amino acid ABC tra
30	27	100.0	305	2	H72289	hypothetical prote
31	27	100.0	306	2	A75316	hypothetical prote
32	27	100.0	318	2	D83725	methanol dehydroge
33	27	100.0	320	2	C69791	methanol dehydroge
34	27	100.0	320	2	G82850	fimbrial adhesin p
35	27	100.0	324	2	T18763	hypothetical prote
36	27	100.0	324	2	A31920	collagen sqt-1 pre
37	27	100.0	332	2	T47703	Ca-dependent solut
38	27	100.0	334	2	C70712	hypothetical prote
39	27	100.0	337	2	T34964	probable regulator
40	27	100.0	344	1	KMBE31	38K protein - huma
41	27	100.0	358	2	T45934	hypothetical prote
42	27	100.0	362	2	JC5386	steroidogenic acut
43	27	100.0	367	2	E69072	hypothetical prote
44	27	100.0	369	2	A75166	hypothetical prote
45	27	100.0	377	2	T27010	hypothetical prote

ALIGNMENTS

```
RESULT  1
T46593
phytoene dehydrogenase [imported] - Mycobacterium marinum (fragment)
C:Species: Mycobacterium marinum
C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 21-Jul-2000
C:Accession: T46593
R:Ramakrishnan, L.; Tran, H.T.; Federspiel, N.A.; Falkow, S.
J. Bacteriol. 179, 5862-5868; 1997
A:Title: A crtB homolog essential for photochromogenicity in Mycobacterium marinum: i
A:Reference number: 223096; MUID:97440138
A:Accession: T46593
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-38 <RAM>
A:Cross-references: EMBL:U92075; NID:g1928930; PIDN:AAB71427.1; PID:g1928931
A:Experimental source: strain M
C:Genetics:
A:Gene: crtI

Query Match             100.0%;  Score 27;  DB 2;  Length 38;
Best Local Similarity   100.0%;  Pred. No. 34;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VPGVG 5
              |||||
Db          1 VPGVG 5


RESULT  2
I45885
elastin - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change 13-Aug-1999
C:Accession: I45885
R:Rosenbloom, J.
Lab. Invest. 51, 605-623, 1984
A:Title: Biology of disease: Elastin: Relation of protein and gene structure to disea
A:Reference number: I45885; MUID:85059254
A:Accession: I45885
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-76 <ROS>
A:Cross-references: GB:M31891; NID:g163008; PIDN:AAA96416.1; PID:g552319
C:Genetics:
A:Introns: 20/1; 58/1
C:Superfamily: elastin

Query Match             100.0%;  Score 27;  DB 2;  Length 76;
Best Local Similarity   100.0%;  Pred. No. 68;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

Qy   1 VPGVG 5
          |||||
Db   34 VPGVG 38

RESULT 3
C40595
hypothetical protein (mutB 3' region) - Streptomyces cinnamonensis (fragment)
C;Species: Streptomyces cinnamonensis
C;Date: 03-May-1994 #sequence_revision 24-Feb-1995
C;Accession: C40595
R;Birch, A.; Leiser, A.; Robinson, J.A.
J. Bacteriol. 175, 3511-3519, 1993
A;Title: Cloning, sequencing, and expression of the gene encoding methylmalonyl-coenzyme
A;Reference number: A40595; MUID:93273720
A;Status: preliminary
A;Accession: C40595
A;Molecule type: DNA
A;Residues: 1-93 <BIR>
A;Cross-references: GB:L10064
C;Genetics:
A;Start codon: GTG

Query Match   100.0%; Score 27; DB 2; Length 93;
Best Local Similarity   100.0%; Pred. No. 82;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   41 VPGVG 45

RESULT 4
A75586
hydrogenase expression/formation HypA-related protein - Deinococcus radiodurans (strain
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 17-Mar-2000
C;Accession: A75586
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Ma
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Me
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: A75586
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-120 <WHI>
A;Cross-references: GB:AE001863; GB:AE001825; NID:g5460670; PIDN:AAF12463.1; PID:g646075
experimental source: strain R1
Genetics:
A;Gene: DRA0316
A;Map position: 2
C;Superfamily: hydrogenase accessory protein

Query Match   100.0%; Score 27; DB 2; Length 120;
Best Local Similarity   100.0%; Pred. No. 1.1e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   67 VPGVG 71

RESULT 5
A72712
hypothetical protein APE1112 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C;Accession: A72712

R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, Aero
A;Reference number: A72450; MUID:99310339
A;Accession: A72712
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <KAW>
A;Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA80097.1; PID:d1043883; PID:g
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1112
C;Superfamily: Aeropyrum pernix hypothetical protein APE1112

Query Match   100.0%; Score 27; DB 2; Length 127;
Best Local Similarity   100.0%; Pred. No. 1.1e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   19 VPGVG 23

RESULT 6
C70680
hypothetical protein Rv2437 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: C70680
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajendream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, R.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: C70680
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-139 <COL>
A;Cross-references: GB:Z81451; GB:AL123456; NID:g3261662; PIDN:CAB03782.1; PID:e28049
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv2437

Query Match   100.0%; Score 27; DB 2; Length 139;
Best Local Similarity   100.0%; Pred. No. 1.2e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   132 VPGVG 136

RESULT 7
B83664
hypothetical protein BH0114 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C;Accession: B83664
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20263314
A;Accession: B83664
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-170 <STO>
A;Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BAB03833.1; GSPDB:G
A;Experimental source: strain C-125

C:Genetics:
A:Gene: BH0114

Query Match　　　　　100.0%;　Score 27; DB 2; Length 170;
Best Local Similarity　100.0%;　Pred. No. 1.5e+02;
Matches　5;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 VPGVG 5
　　　　|||||
Db　58 VPGVG 62

RESULT 8
T05925
hypothetical protein - barley
C:Species: Hordeum vulgare (barley)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 08-Oct-1999
R:Hess, W.R.; Golz, R.R.; Boerner, T.
A:Title: Analysis of randomly selected cDNAs reveals the expression of stress- and defe
Plant Sci. 133, 191-201, 1998
A:Reference number: 215411
A:Accession: T05925
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-181 <HSS>
A:Cross-references: EMBL:AJ222779; NID:e1203989; PIDN:CAA10984.1; PID:e1203990
A:Experimental source: cv. Haisa, leaf

Query Match　　　　　100.0%;　Score 27; DB 2; Length 181;
Best Local Similarity　100.0%;　Pred. No. 1.6e+02;
Matches　5;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　　1 VPGVG 5
　　　　　|||||
Db　120 VPGVG 124

RESULT 9
G64359
adenylate kinase (EC 2.7.4.3) MJ0479 [similarity] - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
rson, J.D.; Sadow P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
eich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock K.G.; Merrick, J.M.; Glodek, A.;
seiece 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999
A:Accession: G64359
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-195 <BUL>
A:Cross-references: GB:U67498; GB:U67117; NID:g1591180; PIDN:AAB98470.1; PID:g1591182; T
C:Genetics:
A:Map position: FOR421829-422416
A:Start codon: GTG
C:Superfamily: Sulfolobus adenylate kinase
C:Keywords: phosphotransferase

Query Match　　　　　100.0%;　Score 27; DB 2; Length 195;
Best Local Similarity　100.0%;　Pred. No. 1.7e+02;
Matches　5;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 VPGVG 5
　　　　|||||
Db　14 VPGVG 18

RESULT 10
E69534
DNA polymerase, bacteriophage-type homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 08-Oct-1999
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343
A:Accession: E69534
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-199 <KLE>
A:Cross-references: GB:AE000782; NID:g2689270; PIDN:AAB88977.1; PID:g264
C:Superfamily: Archaeoglobus Probable DNA-polymerase

Query Match　　　　　100.0%;　Score 27; DB 2; Length 199;
Best Local Similarity　100.0%;　Pred. No. 1.7e+02;
Matches　5;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 VPGVG 5
　　　　|||||
Db　27 VPGVG 31

RESULT 11
E72695
adenylate kinase (EC 2.7.4.3) APE0981 [similarity] - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
R:Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funabashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A:Reference number: A72450; MUID:99310339
A:Accession: E72695
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <KAW>
A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79965.1; PID:g5104650
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0981
C:Superfamily: Sulfolobus adenylate kinase
C:Keywords: phosphotransferase

Query Match　　　　　100.0%;　Score 27; DB 2; Length 204;
Best Local Similarity　100.0%;　Pred. No. 1.8e+02;
Matches　5;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 VPGVG 5
　　　　|||||
Db　15 VPGVG 19

RESULT 12
D70544
probable amidotransferase - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: D70544
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-206 <COL>
A;Cross-references: GB:z95586; GB:AL123456; NID:g3261785; GB:AL123456; NID:g3261785; PID:g2117235
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: hisH
C;Superfamily: amidotransferase hisH; trpG homology

Query Match　　　　100.0%;　Score 27;　DB 2;　Length 206;
Best Local Similarity　100.0%;　Pred. No. 1.8e+02;
Matches　5;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 VPGVG 5
　　　　|||||
Db　47 VPGVG 51

RESULT 13
T45249
probable amidotransferase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C;Accession: T45249
R;James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z22949
A;Accession: T45249
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-206 <JAM>
A;Cross-references: EMBL:AL049913; PIDN:CAB43169.1
A;Experimental source: cosmid B1610
C;Genetics:
A;Note: hisH
C;Superfamily: amidotransferase hisH; trpG homology

Query Match　　　　100.0%;　Score 27;　DB 2;　Length 206;
Best Local Similarity　100.0%;　Pred. No. 1.8e+02;
Matches　5;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 VPGVG 5
　　　　|||||
Db　47 VPGVG 51

RESULT 14
T36699
probable two-component response regulator - Streptomyces coelicolor (fragment)
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000
C;Accession: T36699
R;Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A;Reference number: Z21597
A;Accession: T36699
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-213 <MUR>
A;Cross-references: EMBL:AL049731; PIDN:CAB41738.1; GSPDB:GN00070; SCOEB3:SCH66.11c
C;Genetics:
A;Gene: SCOEB3:SCH66.11c
C;Superfamily: ompR protein; response regulator homology

Query Match　　　　100.0%;　Score 27;　DB 2;　Length 213;
Best Local Similarity　100.0%;　Pred. No. 1.9e+02;
Matches　5;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 VPGVG 5
　　　　|||||
Db　192 VPGVG 196

RESULT 15
T23593
hypothetical protein K10H10.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2000
C;Accession: T23593
submitted to the EMBL Data Library, December 1996
A;Reference number: 219766
R;Percy, C.
A;Accession: T23593
A;Molecule type: DNA
A;Residues: 1-214 <WIL>
A;Cross-references: EMBL:Z83236; PIDN:CAB05780.1; GSPDB:GN00020; CESP:K10H10.4
A;Experimental source: clone K10H10
C;Genetics:
A;Gene: CESP:K10H10.4
A;Map position: 2
A;Introns: 117/3
C;Superfamily: Caenorhabditis elegans hypothetical protein K10H10.4

Query Match　　　　100.0%;　Score 27;　DB 2;　Length 214;
Best Local Similarity　100.0%;　Pred. No. 1.9e+02;
Matches　5;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　　1 VPGVG 5
　　　　|||||
Db　198 VPGVG 202

Search completed: April 24, 2001, 16:42:03
Job time: 468 sec

Tue Apr 24 16:55:38 2001

```
OM protein - protein search, using sw model

              GenCore version 4.5
         Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:        April 24, 2001, 16:40:35 ; Search time 125.5 Seconds
                                          (without alignments)
                                          4.670 Million cell updates/sec

Title:         US-09-340-736-6
Perfect score: 27
Sequence:      1 VPGVG 5

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters:     374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     SPTREMBL_15:*
               1: sp_archea:*
               2: sp_bacteria:*
               3: sp_fungi:*
               4: sp_human:*
               5: sp_invertebrate:*
               6: sp_mammal:*
               7: sp_mhc:*
               8: sp_organelle:*
               9: sp_phage:*
              10: sp_plant:*
              11: sp_rodent:*
              12: sp_unclassified:*
              13: sp_vertebrate:*
              14: sp_virus:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	DB	Length	ID	Description
1	27	100.0	2	38	O05422	O05422 mycobacteri
2	27	100.0	6	76	Q28100	Q28100 bos taurus
3	27	100.0	2	120	Q9RYJ6	Q9ryj6 deinococcus
4	27	100.0	2	127	Q9PD00	Q9pd00 aeropyrum p
5	27	100.0	6	130	Q18832	O18832 equus caba l
6	27	100.0	2	139	P71912	P71912 mycobacteri
7	27	100.0	2	141	O66301	O66301 unidentifie
8	27	100.0	2	141	O66330	O66330 unidentifie
9	27	100.0	2	141	O66336	O66336 unidentifie
10	27	100.0	2	141	O66364	O66364 unidentifie
11	27	100.0	2	141	O66365	O66365 unidentifie
12	27	100.0	2	141	O66367	O66367 unidentifie
13	27	100.0	5	143	Q9N2X5	Q9N2X5 caenorhabdi
14	27	100.0	10	152	Q9XIR0	Q9xir0 arabidopsis
15	27	100.0	5	157	Q9V506	Q9v506 drosophila
16	27	100.0	13	159	Q9PVQ0	Q9pvq0 xenopus lae
17	27	100.0	2	170	Q9KGF1	Q9kgf1 bacillus ha
18	27	100.0	11	176	Q9JU93	Q9ju93 mus musculu
19	27	100.0	4	181	Q9UJK2	Q9ujk2 homo sapien
20	27	100.0	10	183	Q9XGW3	Q9xgw3 oryza sativ
21	27	100.0	3	189	Q9US43	Q9us43 schizosacch
22	27	100.0	1	199	O28007	O28007 archaeoglob
23	27	100.0	2	206	Q9X7C0	Q9x7c0 mycobacteri
24	27	100.0	2	206	O06589	O06589 mycobacteri
25	27	100.0	2	211	O69043	O69043 corynebacte
26	27	100.0	2	213	Q9X943	Q9x943 streptomyce
27	27	100.0	2	214	O45681	O45681 caenorhabdi
28	27	100.0	9	223	O9T1I1	O9t1i1 lactobacill
29	27	100.0	4	228	Q9NV42	Q9NV42 homo sapien
30	27	100.0	1	231	Q9V9O6	Q9v9o6 aeropyrum p
31	27	100.0	2	234	O31379	O31379 cyprinus ca
32	27	100.0	7	234	O31380	O31380 cyprinus ca
33	27	100.0	2	237	Q9RDA0	Q9rda0 streptomyce
34	27	100.0	5	237	O9VCU6	O9vcu6 drosophila
35	27	100.0	2	246	Q9SIN7	Q9sin7 streptomyce
36	27	100.0	4	258	Q9UMF5	Q9umf5 homo sapien
37	27	100.0	2	265	Q9KI91	Q9ki91 bacillus an
38	27	100.0	2	267	Q9P9M3	Q9p9m3 halobacteri
39	27	100.0	2	267	Q9VHH4	Q9vhh4 drosophila
40	27	100.0	2	271	P74665	P74665 synechocyst
41	27	100.0	2	273	Q9KI88	Q9ki88 bacillus ce
42	27	100.0	2	283	Q9RSW3	Q9rsw3 pseudomonas
43	27	100.0	2	285	O51514	O51514 pseudomonas
44	27	100.0	2	285	Q9Z3Y0	Q9z3y0 pseudomonas
45	27	100.0	2	285	Q9X5X8	Q9x5x8 pseudomonas

ALIGNMENTS

```
RESULT   1
O05422
ID     O05422      PRELIMINARY;     PRT;     38 AA.
AC     O05422;
DT     01-JUL-1997 (TrEMBLrel. 04, Created)
DT     01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT     01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE     PHYTOENE DEHYDROGENASE (FRAGMENT)
GN     CRTI.
OS     Mycobacterium marinum.
OC     Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC     Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX     NCBI_TaxID=1781;
RN     [1]
RP     SEQUENCE FROM N.A.
RC     STRAIN=M;
RX     MEDLINE=97440138; PubMed=9294416;
RA     Ramakrishnan L., Tran H.T., Federspiel N.A., Falkow S.;
RT     "A crtB homolog essential for photochromogenicity in Mycobacterium
RT     marinum: isolation, characterization, and gene disruption via
RT     homologous recombination.";
RL     J. Bacteriol. 179:5862-5868(1997).
DR     EMBL; U92075; AAB71427.1; -.
FT     NON_TER      1      1
SQ     SEQUENCE   38 AA;   3986 MW;   6E46332707CDCAB CRC64;

Query Match           100.0%;  Score 27;  DB 2;  Length 38;
Best Local Similarity 100.0%;  Pred. No. 79;
Matches    5; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

Qy         1 VPGVG 5
               |||||
Db         1 VPGVG 5


RESULT   2
Q28100
ID     Q28100      PRELIMINARY;     PRT;     76 AA.
AC     Q28100;
DT     01-NOV-1996 (TrEMBLrel. 01, Created)
DT     01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
```

DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   Elastin (FRAGMENT)
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85059254; PubMed=6150137;
RA   Rosenbloom J.;
RT   "Elastin: relation of protein and gene structure to disease.";
RL   Lab. Invest. 51:605-623(1984).
DR   EMBL; M31891; AAA96416.1; -.
DR   EMBL; M31893; AAA96416.1; JOINED.
DR   EMBL; M31892; AAA96416.1; JOINED.
FT   NON_TER   1   1
SQ   SEQUENCE   76 AA; 6619 MW; E68337BDAEB7B202 CRC64;

Query Match   100.0%; Score 27; DB 6; Length 76;
Best Local Similarity   100.0%; Pred. No. 1.6e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   34 VPGVG 38

RESULT 3
ID   Q9RYJ6
AC   Q9RYJ6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYDROGENASE EXPRESSION/FORMATION HYPA-RELATED PROTEIN.
GN   DRA0316.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome Sequence of the Radioresistant Bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286.1571-1577(1999).
DR   EMBL; AE001863; AAF12463.1; -.
DR   TIGR; DRA0316; -.
DR   INTERPRO; IPR000688; -.
DR   PFAM; PF01155; HypA; -.
SQ   SEQUENCE   120 AA; 12680 MW; F4D6904A72115D4 CRC64;

Query Match   100.0%; Score 27; DB 2; Length 120;
Best Local Similarity   100.0%; Pred. No. 2.6e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   67 VPGVG 71

RESULT 4
Q9YD00

ID   Q9YD00
AC   Q9YD00;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   HYPOTHETICAL 13.6 KDA PROTEIN APE1112.
GN   APE1112.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K., Kubota K., Nakamura Y.,
RA   Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000060; BAA80097.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   127 AA; 13575 MW; F490CD42B1099BD7 CRC64;

Query Match   100.0%; Score 27; DB 1; Length 127;
Best Local Similarity   100.0%; Pred. No. 2.7e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   19 VPGVG 23

RESULT 5
ID   O18832
AC   O18832;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   AGGRECAN CORE PROTEIN (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE PROTEIN)
DE   (CSPCP) (FRAGMENT).
GN   AGC1.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CHONDROCYTES;
RX   MEDLINE=98209637; PubMed=9550267;
RA   Flannery C.R., Little C.B., Caterson B.;
RT   "Molecular cloning and sequence analysis of the aggrecan interglobular
RT   domain from porcine, equine, bovine and ovine cartilage: comparison of
RT   proteinase-susceptible regions and sites of keratan sulfate
RT   substitution.";
RL   Matrix Biol. 16:507-511(1997).
CC   -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC   MATRIX OF CARTILAGENOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC   IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC   HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC   REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC   -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
CC   SIMILARITY).
CC   -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC   TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC   MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC   CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS

```
CC   THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC   AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC   AND G3.
CC   -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC       CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
DR   EMBL; AF019756; AAC68798.1; -.
DR   INTERPRO; IPR000538; -.
DR   PROSITE; PS01241; LINK; PARTIAL.
KW   Glycoprotein; Cartilage; Proteoglycan; Repeat.
FT   NON_TER       1      1
FT   DOMAIN        1      1       LINK 2.
FT   DOMAIN       <1     10       G1-B'.
FT   DOMAIN       <1     10       G1-B.
FT   CARBOHYD     48     48       POTENTIAL.
FT   CARBOHYD     95     95       POTENTIAL.
FT   NON_TER     130    130
SQ   SEQUENCE   130 AA;  13830 MW;  082F79B5AE78B53E CRC64;

Query Match             100.0%;  Score 27;  DB 6;  Length 130;
Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VPGVG 5
              |||||
Db         82 VPGVG 86


RESULT 6
ID   P71912      PRELIMINARY;      PRT;   139 AA.
AC   P71912;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 15.3 KDA PROTEIN.
GN   RV2437 OR MTCY428.09C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393;537-544(1998).
DR   EMBL; Z81451; CAB03782.1; -.
DR   TUBERCULIST; Rv2437; -.
KW   Hypothetical protein.
SQ   SEQUENCE   139 AA;  15315 MW;  84BF115C0E102EC CRC64;

Query Match             100.0%;  Score 27;  DB 2;  Length 139;
Best Local Similarity   100.0%;  Pred. No. 3e+02;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VPGVG 5
              |||||
Db        132 VPGVG 136


RESULT 7
```

```
ID   O66301      PRELIMINARY;      PRT;   141 AA.
AC   O66301;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ALPHA SUBUNIT OF DINITROGENASE REDUCTASE (FE PROTEIN) (FRAGMENT).
GN   NIFH.
OS   unidentified nitrogen-fixing bacteria.
OC   Bacteria.
OX   NCBI_TaxID=34107;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohkuma M., Noda S., Kudo T.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB011855; BAA28392.1; -.
DR   HSSP; P00456; ICP2.
DR   INTERPRO; IPR000392; -.
DR   PFAM; PF00142; fer4_NifH; 1.
DR   PRINTS; PR00091; NITROGNASEII.
DR   PROSITE; PS00692; NIFH_FRXC_2; 1.
DR   PROSITE; PS00746; NIFH_FRXC_1; 1.
FT   NON_TER     141    141
SQ   SEQUENCE   141 AA;  14882 MW;  02454DCD55EDBF40 CRC64;

Query Match             100.0%;  Score 27;  DB 2;  Length 141;
Best Local Similarity   100.0%;  Pred. No. 3.1e+02;  Indels    0;  Gaps    0;
Matches    5;  Conservative    0;  Mismatches    0;

Qy          1 VPGVG 5
              |||||
Db         74 VPGVG 78


RESULT 8
ID   O66330      PRELIMINARY;      PRT;   141 AA.
AC   O66330;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ALPHA SUBUNIT OF DINITROGENASE REDUCTASE (FE PROTEIN) (FRAGMENT).
GN   NIFH.
OS   unidentified nitrogen-fixing bacteria.
OC   Bacteria.
OX   NCBI_TaxID=34107;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohkuma M., Noda S., Kudo T.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB011900; BAA28435.1; -.
DR   HSSP; P00456; ICP2.
DR   INTERPRO; IPR000392; -.
DR   PFAM; PF00142; fer4_NifH; 1.
DR   PRINTS; PR00091; NITROGNASEII.
DR   PROSITE; PS00692; NIFH_FRXC_2; 1.
DR   PROSITE; PS00746; NIFH_FRXC_1; 1.
FT   NON_TER     141    141
SQ   SEQUENCE   141 AA;  14632 MW;  7251716CF85C6752 CRC64;

Query Match             100.0%;  Score 27;  DB 2;  Length 141;
Best Local Similarity   100.0%;  Pred. No. 3.1e+02;  Indels    0;  Gaps    0;
Matches    5;  Conservative    0;  Mismatches    0;

Qy          1 VPGVG 5
              |||||
Db         74 VPGVG 78


RESULT 9
```

O66336
ID   O66336   PRELIMINARY;   PRT;   141 AA.
AC   O66336;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ALPHA SUBUNIT OF DINITROGENASE REDUCTASE (FE PROTEIN) (FRAGMENT).
GN   NIFH.
OS   unidentified nitrogen-fixing bacteria.
OC   Bacteria.
OX   NCBI_TaxID=34107;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohkuma M., Noda S., Kudo T.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB011906; BAA28441.1; -.
DR   HSSP; P00456; 1CP2.
DR   INTERPRO; IPR000392; -.
DR   PFAM; PF00142; fer4_NifH; 1.
DR   PRINTS; PR00091; NITROGNASEII.
DR   PROSITE; PS00692; NIFH_FRXC_2; 1.
DR   PROSITE; PS00746; NIFH_FRXC_1; 1.
FT   NON_TER   1   1
FT   NON_TER   141   141
SQ   SEQUENCE   141 AA; 14910 MW; D2E0514869D28C3 CRC64;

Query Match   100.0%; Score 27; DB 2; Length 141;
Best Local Similarity   100.0%; Pred. No. 3.1e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   74 VPGVG 78

RESULT 10
O66364
ID   O66364   PRELIMINARY;   PRT;   141 AA.
AC   O66364;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ALPHA SUBUNIT OF DINITROGENASE REDUCTASE (FE PROTEIN) (FRAGMENT).
GN   NIFH.
OS   unidentified nitrogen-fixing bacteria.
OC   Bacteria.
OX   NCBI_TaxID=34107;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohkuma M., Noda S., Kudo T.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB011954; BAA28487.1; -.
DR   HSSP; P00456; 1CP2.
DR   INTERPRO; IPR000392; -.
DR   PFAM; PF00142; fer4_NifH; 1.
DR   PROSITE; PS00692; NIFH_FRXC_2; 1.
DR   PROSITE; PS00746; NIFH_FRXC_1; 1.
FT   NON_TER   1   1
FT   NON_TER   141   141
SQ   SEQUENCE   141 AA; 14780 MW; E069FA235F62484B CRC64;

Query Match   100.0%; Score 27; DB 2; Length 141;
Best Local Similarity   100.0%; Pred. No. 3.1e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   74 VPGVG 78

RESULT 11

O66365
ID   O66365   PRELIMINARY;   PRT;   141 AA.
AC   O66365;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ALPHA SUBUNIT OF DINITROGENASE REDUCTASE (FE PROTEIN) (FRAGMENT).
GN   NIFH.
OS   unidentified nitrogen-fixing bacteria.
OC   Bacteria.
OX   NCBI_TaxID=34107;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohkuma M., Noda S., Kudo T.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB011955; BAA28488.1; -.
DR   HSSP; P00456; 1CP2.
DR   INTERPRO; IPR000392; -.
DR   PFAM; PF00142; fer4_NifH; 1.
DR   PROSITE; PS00692; NIFH_FRXC_2; 1.
DR   PROSITE; PS00746; NIFH_FRXC_1; 1.
FT   NON_TER   1   1
FT   NON_TER   141   141
SQ   SEQUENCE   141 AA; 14785 MW; 2BC10A2BF5C05E2E CRC64;

Query Match   100.0%; Score 27; DB 2; Length 141;
Best Local Similarity   100.0%; Pred. No. 3.1e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   74 VPGVG 78

RESULT 12
O66367
ID   O66367   PRELIMINARY;   PRT;   141 AA.
AC   O66367;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ALPHA SUBUNIT OF DINITROGENASE REDUCTASE (FE PROTEIN) (FRAGMENT).
GN   NIFH.
OS   unidentified nitrogen-fixing bacteria.
OC   Bacteria.
OX   NCBI_TaxID=34107;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohkuma M., Noda S., Kudo T.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB011957; BAA28490.1; -.
DR   HSSP; P00456; 1CP2.
DR   INTERPRO; IPR000392; -.
DR   PFAM; PF00142; fer4_NifH; 1.
DR   PROSITE; PS00692; NIFH_FRXC_2; 1.
DR   PROSITE; PS00746; NIFH_FRXC_1; 1.
FT   NON_TER   1   1
FT   NON_TER   141   141
SQ   SEQUENCE   141 AA; 14505 MW; 7A2B008BB564C29C CRC64;

Query Match   100.0%; Score 27; DB 2; Length 141;
Best Local Similarity   100.0%; Pred. No. 3.1e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   74 VPGVG 78

RESULT 13
Q9N2X5

```
ID   Q9N2X5;       PRELIMINARY;       PRT;    143 AA.
AC   Q9N2X5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   Y92C33.B PROTEIN.
GN   Y92C33.B
OS   Caenorhabditis elegans
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99096613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC024875; AAF50885.1; -.
SQ   SEQUENCE 143 AA; 16000 MW; CDFD9C94F94A00D7 CRC64;

Query Match         100.0%;  Score 27;  DB 5;  Length 143;
Best Local Similarity 100.0%;  Pred. No. 3.1e+02;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 VPGVG 5
           |||||
Db    95 VPGVG 99


RESULT 14
ID   Q9XIRO;       PRELIMINARY;       PRT;    162 AA.
AC   Q9XIRO;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   F13O11.15 PROTEIN.
GN   F13O11.15
OS   Arabidopsis thaliana (Mouse-ear cress)
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv COLUMBIA;
RA   Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA   Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA   Gonzales A., Kremenetskaia I., Kim C., Lenz C.A., Li J., Liu S.,
RA   Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA   Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RA   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006193; AAD38259.1; -.
DR   INTERPRO; IPR002048; -.
DR   PFAM; PF00036; efhand; 1.
SQ   SEQUENCE 162 AA; 18098 MW; D9ADA7D742F25618 CRC64;

Query Match         100.0%;  Score 27;  DB 10;  Length 162;
Best Local Similarity 100.0%;  Pred. No. 3.5e+02;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 VPGVG 5
           |||||
Db   131 VPGVG 135


RESULT 15
ID   Q9V506;       PRELIMINARY;       PRT;    167 AA.
AC   Q9V506;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG13749 PROTEIN.
GN   CG13749;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKLEY;
RX   MEDLINE=20196005; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003835; AAF59017.1; -.
DR   FLYBASE; FBgn0033353; CG13749.
SQ   SEQUENCE 167 AA; 18196 MW; E147CAC7C422F50D CRC64;

Query Match         100.0%;  Score 27;  DB 5;  Length 167;
Best Local Similarity 100.0%;  Pred. No. 3.6e+02;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 VPGVG 5
           |||||
Db    52 VPGVG 56
```

```
              GenCore version 4.5
         Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:      April 24, 2001, 16:38:19 ; Search time 115.25 seconds
                                 (without alignments)
                             1.984 Million cell updates/sec

Title:         US-09-340-736-7
Perfect score: 23
Sequence:      1 VPGG 4

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      .390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters:      390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    A_Geneseq_0401:*
          1:  /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
          2:  /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
          3:  /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
          4:  /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
          5:  /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
          6:  /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
          7:  /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
          8:  /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
          9:  /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
         10:  /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
         11:  /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
         12:  /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
         13:  /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
         14:  /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
         15:  /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
         16:  /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
         17:  /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
         18:  /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
         19:  /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
         20:  /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
         21:  /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
         22:  /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

                              SUMMARIES

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Result NO.	Score	Query Match	Length	DB	ID	Description
1	23	100.0	4	7	P60729	Synthetic repeatin
2	23	100.0	4	16	R65231	Fibrinogen/beta-ca
3	23	100.0	4	18	W18264	Transglutaminase c
4	23	100.0	4	20	Y31685	Tetrapeptide used
5	23	100.0	4	21	B19213	Elastin fragment u
6	23	100.0	4	21	Y80336	Repeat elastin tet
7	23	100.0	7	22	B63971	Elastin repeating
8	23	100.0	7	17	W07157	Synthetic peptide
9	23	100.0	7	19	W59326	Non-polio enterovi
10	23	100.0	7	20	Y50084	Coxsackievirus A v
11	23	100.0	9	17	W07039	Synthetic peptide

```
         12   23   100.0         10  10  P91309   Sequence of beta-t
         13   23   100.0         10  17  W02011   Peptide derived fr
         14   23   100.0         10  19  W41471   Fragment of nemato
         15   23   100.0         11  20  Y31331   B. subtilis surfac
         16   23   100.0         12  17  W01993   Peptide derived fr
         17   23   100.0         13  17  R98232   Tau protein phosph
         18   23   100.0         13  19  W41872   Peptide used in ra
         19   23   100.0         13  21  B21965   Phosphorylated tau
         20   23   100.0         14  18  W10311   Gamma-subtype reti
         21   23   100.0         16  14  R37957   Peptide SP14 (huma
         22   23   100.0         17  17  R98230   Tau protein phosph
         23   23   100.0         19  15  R47084   Eukaryotic folding
         24   23   100.0         19  21  Y98289   Alpha D peptide de
         25   23   100.0         19  21  Y98290   Alpha D peptide de
         26   23   100.0         20  21  W90974   Human periplasmic
         27   23   100.0         24  12  R14997   Part of elastomeri
         28   23   100.0         24  12  R14983   Part of elastomeri
         29   23   100.0         24  12  R14990   Part of elastomeri
         30   23   100.0         24  12  R15025   Part of elastomeri
         31   23   100.0         24  21  R15032   Yeast GAL1 protein
         32   23   100.0         24  21  Y81863   Part of elastomeri
         33   23   100.0         26  12  R15016   Part of elastomeri
         34   23   100.0         26  12  R15011   Part of elastomeri
         35   23   100.0         26  12  R14955   Part of elastomeri
         36   23   100.0         26  12  R14962   Part of elastomeri
         37   23   100.0         26  12  R14969   Part of elastomeri
         38   23   100.0         26  12  R15004   Part of elastomeri
         39   23   100.0         26  12  R14976   Part of elastomeri
         40   23   100.0         27   9  P82594   Upstream open read
         41   23   100.0         29  21  B45205   Human secreted pro
         42   23   100.0         30  20  Y12930   Amino acid sequenc
         43   23   100.0         31  21  V55106   Tau protein fragme
         44   23   100.0         35  19  W75127   Human secreted pro
         45   23   100.0         38  21  G59824   Arabidopsis thalia


                                      ALIGNMENTS

RESULT    1
P60729
ID   P60729 standard; protein; 4 AA.
XX
AC   P60729;
XX
DT   01-AUG-1991  (first entry)
XX
DE   Synthetic repeating unit.
XX
KW   Elastomeric copolymer; prosthesis; repeating unit.
XX
PN   US4589882-A.
XX
PD   20-MAY-1986.
XX
PF   19-SEP-1983;   83US-0533524.
XX
PR   19-SEP-1983;   83US-0533524.
XX
PA   (URRY/) URRY D W.
XX
PI   Urry DW;
XX
DR   WPI; 1986-149845/23.
XX
PT   Synthetic elastomeric copolymers - useful as prostheses for
PT   repair of ligaments, tendons and blood vessel walls.
XX
PS   Disclosure; page 9; 11pp; English.
XX
CC   The repeating unit can be used in the prodn. of a synthetic elastic
CC   copolymer. The copolymer contains a beta turn and the repeating
CC   unit contains hydrophobic amino acid and glycine residues and a cross-
```

CC linking component. The copolymer is useful in prosthetic systems, for
CC repairing a natural elastic system. It is so functionalised so as to
CC provide reactive gps. which can become covalently crosslinked by
CC tissue enzymes to newly synthesised connective tissue protein.
CC The copolymer comprises 15 units of VPGVG, 5 units of the block
CC unit and 1 unit of AAAALATGA. It has a mol. wt. of 80000.
XX
SQ Sequence 4 AA;

Qy 1 VPGG 4
       ||||
Db 1 vpgg 4

Query Match   100.0%; Score 23; DB 7; Length 4;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 2
R65231 R65231 standard: peptide; 4 AA.
XX
ID R65231;
XX
DT 12-OCT-1995 (first entry)
XX
DE Fibrinogen/beta-casein transglutaminase cross-linkable copolymer.
XX
KW Transglutaminase cross-linkable copolymer; fibrinogen; beta-casein;
KW biocompatible; bioadhesive polymers; wound repair; prosthetics;
KW bone and soft tissue matrices; controlled drug release carriers.
XX
OS Synthetic.
XX
PN WO9505396-A.
XX
PD 23-FEB-1995.
XX
PF 05-AUG-1994; 94WO-US08754.
PR 13-AUG-1993; 93US-0106509.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Busby SJ; Labroo VM;
XX
DR WPI; 1995-098722/13.
XX
PT New peptide(s) based on fibrinogen and beta-casein -
PT cross-linkable by trans:glutaminase, used for preparing
PT biocompatible, bio:adhesive polymers
XX
PS Claim 16; Page 48; 58pp; English.
XX
CC A biocompatible, bioadhesive, transglutaminase cross-linkable
CC copolymer comprises a first polypeptide monomer from 13-120 amino
CC acids containing a segment of the formula S1-Y-S2 cross-linkable
CC by a transglutaminase, where S1 = R65218, S2 = R65219 and Y = a
CC spacer of 0-7 amino acids pref. R65220/21/22 or R65223, and
CC a second polypeptide monomer selected from R65228-R65231 (the
CC first monomer can also be R65227). The copolymer can be used to
CC produce tissue adhesives, wound repair formulations, rigid
CC prosthetics, matrices for the replacement of bone and soft tissue
CC structures and carriers for controlled drug release compns.
XX
SQ Sequence 4 AA;

Query Match   100.0%; Score 23; DB 15; Length 4;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 1 VPGG 4
       ||||
Db 1 vpgg 4

RESULT 3
WI8264 ID WI8264 standard: peptide; 4 AA.
XX
AC WI8264;
XX
DT 29-AUG-1997 (first entry)
XX
DE Transglutaminase cross-linkable polypeptide elastomeric peptide.
XX
KW Elastomeric; homopolymer; copolymer; tissue sealant; skin graft;
KW burn.
XX
OS Synthetic.
XX
PN WO9640780-A1.
XX
PD 19-DEC-1996.
XX
PF 31-MAY-1996; 96WO-US08269.
PR 07-JUN-1995; 95US-0483236.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Busby SJ; Labroo VM;
XX
DR WPI; 1997-052237/05.
XX
PT Trans:glutaminase cross-linkable peptide(s) - used in the mfr. of
PT biocompatible, bio:adhesive tissue sealant and wound healing
PT preparations.
XX
PS Claim 6; Page 48; 55pp; English.
XX
CC Novel polypeptides optionally having one or both of the amino-terminus
CC and carboxy-terminus flanked by an elastomeric peptide, which is cross-
CC linkable by a transglutaminase comprises a segment of formula:
CC S1-Y-S2. The present sequence represents a specifically claimed
CC example of an elastomeric peptide. The homo- and copolymers produced
CC are useful in tissue sealant and wound healing formulations. Tissue
CC sealants are useful in skin grafting for burn victims and for sealing
CC surgical and other wounds.
XX
SQ Sequence 4 AA;

Query Match   100.0%; Score 23; DB 18; Length 4;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 1 VPGG 4
       ||||
Db 1 vpgg 4

RESULT 4
Y31685 ID Y31685 standard; Peptide; 4 AA.
XX
AC Y31685;
XX
DT 22-NOV-1999 (first entry)
XX
DE Tetrapeptide used in novel elastomer polymers.
XX
KW Elastomer; bioelastomer; polymer; tissue augmentation;
KW tissue restoration; tissue reconstruction; tissue repair; implant.

XX
OS Synthetic.
XX
PN WO9943271-A1.
XX
PD 02-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04440.
XX
PR 29-MAY-1998; 98US-0087155.
PR 27-FEB-1998; 98US-0076297.
XX
PA (BIOE-) BIOELASTICS RES LTD.
XX
PI Glazer PA, Parker TM, Urry DW;
XX
DR WPI; 1999-540487/45.
XX
PT Augmentation or restoration of mammalian tissue by injecting
PT solution of peptide polymer, used for soft or hard tissue
PT reconstruction, especially of intervertebral disks
XX
PS Claim 8; Page 75; 133pp; English.
XX
CC The present sequence represents an example of a tetrapeptide that
CC is used in novel bioelastic polymers. The invention provides a
CC method of tissue augmentation by injecting a polymer comprising
CC repeating peptide monomeric units selected from nonapeptide,
CC pentapeptide and tetrapeptide monomeric units, where the monomeric
CC units form a series of beta-turns separated by dynamic bridging
CC segments. The polymer has an inverse temperature transition value
CC that is less than the tissue temp, and is injected in water
CC solution at concervate concentration. The polymer can be injected
CC at periurethral or subdermal sites (for treatment of urinary
CC incontinence or for cosmetic purposes), or into hard or soft
CC tissue e.g. for repair of traumatic injury. A specific
CC application is restoration of intervertebral discs.
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 23; DB 20; Length 4;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPGG 4
   ||||
Db 1 vpgg 4

RESULT 5
B19213
ID B19213 standard; Peptide: 4 AA.
XX
AC B19213;
XX
DT 19-FEB-2001 (first entry)
XX
DE Elastin fragment used to make hybrid bay scallop abductions.
XX
KW Bay scallop; abductin; chemomechanical transduction; drug delivery;
KW inverse temperature transition; water soluble drug; biomaterial;
KW fabric; organ prosthesis.
XX
OS Synthetic.
XX
PN US6127166-A.
XX
PD 03-OCT-2000.
XX
PF 03-NOV-1997; 97US-0963168.
XX
PR 03-NOV-1997; 97US-0963163.

XX
PA (BAYL/) BAYLEY H.
PA (CAOQ/) CAO Q.
PA (WANG/) WANG Y.
XX
PI Bayley H, Cao Q, Wang Y;
XX
DR WPI; 2000-611057/58.
XX
PT Abductin nucleic acid molecules, useful for expressing abductin
PT polypeptides which are used in the manufacture of drug delivery
PT vehicles for administering water soluble drugs -
XX
PS Disclosure; Column 14; 30pp; English.
XX
CC The present sequence is used to make hybrid bay scallop abductin
CC polypeptide sequences. Abductin is capable of chemomechanical
CC transduction or inverse temperature transition. Abductin polypeptides
CC contain glycine-rich sequences. Abductin polypeptides are useful in
CC the manufacture of drug delivery vehicles for administering water
CC soluble drugs. The abductin polypeptides and their derivatives are
CC also useful in the manufacture of broad range of biomaterials ranging
CC from light-weight durable fabric for clothing to matrices useful for
CC human tissue and organ prosthesis.
XX
SQ Sequence 4 AA;

Query Match 100.0%; Score 23; DB 21; Length 4;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VPGG 4
   ||||
Db 1 vpgg 4

RESULT 6
Y80336
ID Y80336 standard; peptide; 4 AA.
XX
AC Y80336;
XX
DT 24-MAY-2000 (first entry)
XX
DE Repeat elastin tetrapeptide from tropoelastin.
XX
KW Overexpression; bioelastic polypeptide; pentapeptide; tetrapeptide;
KW hexapeptide; nonapeptide; repeat unit; inclusion body; elastin;
KW inverse temperature transition; tropoelastin.
XX
OS Bos taurus.
XX
PN US6004782-A.
XX
PD 21-DEC-1999.
XX
PF 13-OCT-1995; 95US-0542051.
XX
PR 14-APR-1995; 95US-0423642.
XX
PA (BIOE-) BIOELASTICS RES LTD.
PA (UABR-) UAB RES FOUND.
XX
PI Xu J, Daniell H, McPherson DT, Urry DW;
XX
DR WPI; 2000-125738/11.
XX
PT Overexpression of bioelastic polypeptides in prokaryotic cells,
PT exhibiting an inverse temperature transition -
XX
PS Disclosure; Column 2; 32pp; English.
XX

CC   The invention relates to overexpressing a bioelastic polypeptide in
CC   a prokaryotic host cell by introducing into the host cell a vector
CC   containing a promoter operably linked to a nucleic acid encoding a
CC   bioelastic polypeptide, and growing the host cell for expressing
CC   the polypeptide. The bioelastic polypeptide especially comprises
CC   pentapeptide, tetrapeptide, hexapeptide, or nonapeptide repeat units.
CC   The host cell is grown such that the overexpressed peptide is
CC   produced in inclusion bodies, where the volume of inclusion bodies
CC   comprises 40-90% (v/v) of the total cellular volume of the host cell
CC   The method can be used to produce bioelastic polypeptides or polymers
CC   that exhibit an inverse temperature transition. This sequence
CC   represents the natural elastin tetrapeptide repeat from tropoelastin.
XX
SQ   Sequence   4 AA;

Query Match   100.0%; Score 23; DB 21; Length 4;
Best Local Similarity   100.0%; Pred. No. 3.2e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
               ||||
Db   1 vpgg 4

RESULT 7
B63971
ID   B63971 standard; Peptide; 4 AA.
XX
AC   B63971;
XX
DT   19-MAR-2001 (first entry)
XX
DE   Elastin repeating unit peptide sequence SEQ ID 1.
XX
KW   Proteinaceous polymer; repeat unit; structural polymer; coating; film;
KW   fibre; membrane; adhesive; emulsion; laminate; keratin; collagen.
XX
OS   Unidentified.
XX
PN   US6140072-A.
XX
PD   31-OCT-2000.
XX
PF   07-JUN-1995; 95US-0475411.
XX
PR   06-NOV-1990; 90US-0609716.
PR   04-NOV-1986; 86US-0927258.
PR   29-OCT-1987; 87US-0114618.
PR   09-NOV-1988; 88US-0269429.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J, Ferrari FA;
XX
DR   WPI; 2001-048958/06.
XX
PT   New DNA encoding a polymer with strands of repeating units of natural
PT   protein joined by intervening oligopeptide for producing high molecular
PT   weight polymers of amino acids
XX
PS   Claim 3; Column 143; 73pp; English.
XX
CC   This invention relates to DNA encoding a proteinaceous polymer. The
CC   polymer comprises strands of repeating units of a natural protein capable
CC   of assembling into aligned structures, with at least 2 strands joined by
CC   an intervening oligopeptide other than the repeating units. The
CC   intervening oligopeptide is unaligned and the polymer has individual
CC   strands of the same or different repeating units. The DNA is useful for
CC   producing high molecular weight polymers of amino acids based on
CC   biologically and chemically active structural polymers. These polymers
CC   may be used to provide a variety of structures for different purposes,
CC   and to produce articles including coatings, or other (non)structural

CC   components, e.g. fibres, films, membranes, adhesives or emulsions, or
CC   with other compounds and/or composites to form composites or laminates.
CC   Peptide sequences B63971-B63991 represent monomer sequences which can be
CC   used in the polymers of the invention. Oligonucleotide sequences
CC   F23370 - F23386 and amino acid sequences B63992 - B64002 are used in the
CC   construction of SLP and PCB-SLP polymers. Oligonucleotide sequences
CC   F23387 - F23397 and amino acid sequences B64003 - B64008 are used in the
CC   construction of CLP (collagen like protein) polymers. Oligonucleotide
CC   sequences F23398 - F23409 and amino acid sequences B64009 - B64014 are
CC   used in the construction of keratin polymers. Proteins and peptides
CC   represented by sequences B64015 - B64049 are examples of polymers of the
CC   invention.
XX
SQ   Sequence   4 AA;

Query Match   100.0%; Score 23; DB 22; Length 4;
Best Local Similarity   100.0%; Pred. No. 3.2e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
               ||||
Db   1 vpgg 4

RESULT 8
W07157
ID   W07157 standard; peptide; 7 AA.
XX
AC   W07157;
XX
DT   24-JAN-1997 (first entry)
XX
DE   Synthetic peptide used in GalNac-transferase activity SPA.
XX
KW   SPA; scintillation proximity assay; antigen; bead coating; capture;
KW   antibody; N-acetyl galactosamine transferase; GalNac transferase;
KW   activity; enzyme; O-linked glycosylation.
XX
OS   Synthetic.
XX
PN   WO9615258-A1.
XX
PD   23-MAY-1996.
XX
PF   08-NOV-1995;   95WO-US13483.
XX
PR   16-NOV-1994;   94US-0340283.
XX
PA   (UPJO) UPJOHN CO.
XX
PI   Elhammer AP;
XX
DR   WPI; 1996-268220/27.
XX
PT   Scintillation proximity assay for N-acetyl-galactosaminyl activity
PT   - esp. for large scale screening of cpds. for their effect on enzyme
PT   activity
XX
PS   Claim 14; Page 17; 29pp; English.
XX
CC   W06985-W07180 are antigenic peptides derived from either the
CC   product of the human c-myc oncogene or the FLAG peptide (DYKDDDK).
CC   The peptides are useful for coating beads used in a scintillation
CC   proximity assay for N-acetylgalactosamine (GalNac)-transferase (GNT)
CC   activity. The assay involves fewer steps than known assays and is
CC   quicker, producing excellent signal-to-noise ratios. The
CC   assay is capable of screening large numbers of cpds. for their
CC   ability to affect GNT activity and is thus useful for identifying
CC   inhibitors and promoters of glycosylation (in partic. O-linked
CC   glycosylation).
XX
SQ   Sequence   7 AA;

```
Query Match      100.0%;  Score 23;  DB 17;  Length 7;
Best Local Similarity   100.0%;  Pred. No. 3.2e+05;
Matches   4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 VPGG 4
        ||||
Db    4 vpgg 7
```

RESULT 9
W59326
ID W59326 standard; peptide; 7 AA.
XX
AC W59326;
XX
DT 24-SEP-1998 (first entry)
XX
DE Non-polio enterovirus peptide fragment 61S.
XX
KW Non-polio enterovirus; NPEV; enteroviral disease; aseptic meningitis;
KW vaccination.
XX
OS Enterovirus sp.
XX
PN WO9814611-A2.
XX
PD 09-APR-1998.
XX
PF 01-OCT-1997; 97WO-US17734.
PR 02-OCT-1996; 96US-0027353.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Kilpatrick D;
PT WPI; 1998-240106/21.
XX
PT Identifying non-polio enteroviruses - using primers which hybridise
PT to sense and antisense strands that encode conserved non-polio
PT enterovirus peptide sequences
XX
PS Claim 4; Page 23; 47pp; English.
XX
CC The peptide sequences W59298-W59344 are amplified by primers to detect
CC the presence of a non-polio enterovirus (NPEV)in a sample. The primers
CC and assays are used to detect NPEVs in a sample, to serotype these
CC viruses, to diagnose enteroviral diseases and medical conditions, and to
CC correlate (or disprove a correlation between) specific symptoms or
CC combinations of symptoms with the presence of a particular enterovirus.
CC They can be used for diseases such as aseptic meningitis. The detection
CC of NPEV infections and their correlation with medical conditions will
CC make possible vaccines and methods of treatment.
XX
SQ Sequence 7 AA;

```
Query Match      100.0%;  Score 23;  DB 19;  Length 7;
Best Local Similarity   100.0%;  Pred. No. 3.2e+05;
Matches   4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 VPGG 4
        ||||
Db    3 vpgg 5
```

RESULT 10
Y50084
ID Y50084 standard; peptide; 7 AA.
XX
AC Y50084;

XX
DT 19-JAN-2000 (first entry)
XX
DE Coxsackievirus A VP1 conserved epitope 61.
XX
KW Virus; epitope; target; degenerate; PCR; primer; amplification;
KW VP1; nonstructural protein 2A; conserved; base analogue; inosine;
KW predetermined nucleotide; diagnosis; enterovirus; poliovirus.
XX
OS Synthetic.
OS Coxsackievirus.
XX
PN WO9953097-A2.
XX
PD 21-OCT-1999.
XX
PF 06-APR-1999; 99WO-US07513.
PR 15-APR-1998; 98US-0081944.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Kilpatrick DR;
XX
DR WPI; 1999-620444/53.
DR N-PSDB; Z32603.
XX
PT Designing degenerate polymerase chain reaction primers -
XX
PS Example 3; Page 18; 30pp; English.
XX
CC This sequence represents a conserved Coxsackievirus A
CC (serotype A21) VP1 epitope. The invention relates to a novel
CC method for designing degenerate PCR primers (230975-231000, 232601-
CC 232611) for amplifying target polynucleotides. This method comprises
CC identifying uniquely conserved amino acid sequences (e.g, this
CC epitope) in target proteins; synthesising degenerate polynucleotides
CC encoding the conserved sequences; and substituting the synthesised
CC polynucleotides with up to four predetermined nucleotides (e.g.,
CC inosine) at degenerate nucleotide positions. The nucleic acids
CC comprise no more than 7 degenerate positions, have no more
CC than 2 adjacent predetermined nucleotides and the predetermined
CC nucleotides are 3 bases away from the 3' end of the synthesised strand.
CC The degenerate primers are useful for amplifying target polynucleotides
CC by the polymerase chain reaction (PCR). The use of the method of
CC designing degenerate primers useful for the detection of polioviruses
CC in clinical samples is described in US585477. The degenerate primers
CC facilitate PCR amplification of unknown polynucleotides, where the amino
CC acid sequence encoded is known. The primers also allow for the
CC correlation of the subsequent molecular based diagnosis with a
CC serologically derived diagnosis.
XX
SQ Sequence 7 AA;

```
Query Match      100.0%;  Score 23;  DB 20;  Length 7;
Best Local Similarity   100.0%;  Pred. No. 3.2e+05;
Matches   4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 VPGG 4
        ||||
Db    3 vpgg 6
```

RESULT 11
W07039
ID W07039 standard; peptide; 9 AA.
XX
AC W07039;
XX
DT 17-JAN-1997 (first entry)
XX
DE Synthetic peptide used in GalNac-transferase activity SPA.

XX
KW   SPA; scintillation proximity assay; antigen; bead coating; capture;
KW   antibody; N-acetyl galactosamine transferase; GalNac transferase;
KW   activity; enzyme; O-linked glycosylation.
XX
OS   Synthetic.
XX
PN   WO9615258-A1.
XX
PD   23-MAY-1996.
XX
PF   08-NOV-1995;   95WO-US13483.
XX
PR   16-NOV-1994;   94US-0340283.
XX
PA   (UPJO) UPJOHN CO.
XX
PI   Elhammer AP;
XX
DR   WPI; 1996-268220/27.
XX
PS   Claim 14; Page 17; 29pp; English.
XX
CC   Scintillation proximity assay for N-acetyl:galactosaminyl activity
CC   - esp. for large scale screening of cpds. for their effect on enzyme
CC   activity
CC
CC   WO6985-W07180 are antigenic peptides derived from either the
CC   product of the human c-myc oncogene or the FLAG peptide (DYKDDDK).
CC   The peptides are useful for coating beads used in a scintillation
CC   proximity assay for N-acetylgalactosamine (GalNAc) transferase (GNT)
CC   activity. The assay involves fewer steps than known assays and is
CC   quicker, producing excellent signal-to-noise ratios. The
CC   assay is capable of screening large numbers of cpds. for their
CC   ability to affect GNT activity and is thus useful for identifying
CC   inhibitors and promoters of glycosylation (in partic. O-linked
CC   glycosylation).
XX
SQ   Sequence 9 AA;

Query Match   100.0%; Score 23; DB 17; Length 9;
Best Local Similarity 100.0%; Pred. No. 3.2e+05;
Matches 4;   Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 VPGG 4
       ||||
Db   1 vpgg 4

RESULT 12
1309
ID   P91309 standard; Protein; 10 AA.
XX
AC   P91309;
XX
DT   09-MAR-1992 (first entry)
XX
DE   Sequence of beta-turn of a bioelastomeric material.
XX
KW   Bioelastomer; elastomeric material; wound dressing; burn;
KW   artificial veins; arteries; skin; ligament; biodegradable.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 1..3
FT   /note= "May be PGG, GG, G or a covalent
FT   bond"
FT   Misc-difference 8..10
FT   /note= "May be VPG, VP, V or a covalent
FT   bond"
FT   Misc-difference 4..7
FT   /note= "May be absent or present up to 200
FT   times"

XX
PN   WO8910099-A.
XX
PD   02-NOV-1989.
XX
PF   14-APR-1989;   89WO-US01482.
XX
PR   21-APR-1988;   88US-0184407.
XX
PA   (UABR-) UAB-RES FOUNDATION.
XX
PI   Urry DW;
XX
DR   WPI; 1989-339743/46.
XX
PT   Elastomeric polypeptide material - a useful for preventing
PT   adhesion between tissues and wound repair sites
XX
PS   Claim 8; page 86; 93pp; English.
XX
CC   The elastomeric material of the invention comprises a bioelastomer
CC   contg. repeating elastoeric tetrapeptide or pentapeptide units opt.
CC   modified by hexapeptide units, the units consisting of hydrophobic
CC   AA residues and Gly residues and existing in a conformation having
CC   a beta-turn. The beta-turn of the the bioelastomer comprises
CC   a polypeptapeptide unit of VPGVG (see P91307) and/or IPGVG
CC   (see P91308), or a pentapeptide repeating unit of formula XPXXG
CC   (see P91311), or a polytetrapeptide of formula VPGG (see P91309),
CC   IPGG (see P91310), or XPGG (see P91312); and the hexapeptide
CC   repeating unit of formula APGVGV (see P91313).
XX
SQ   Sequence 10 AA;

Query Match   100.0%; Score 23; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 4;   Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 VPGG 4
       ||||
Db   4 vpgg 7

RESULT 13
WO2011
ID   WO2011 standard; peptide; 10 AA.
XX
AC   WO2011;
XX
DT   30-SEP-1996 (first entry)
XX
DE   Peptide derived from hepatitis C virus NS4 protein.
XX
KW   Antigenic peptide; hepatitis C virus; HCV; non-A non-B;
KW   non-structural protein 4; non-reactive; NS4;
XX
OS   Hepatitis C virus.
XX
PN   WO9604300-A1.
XX
PD   15-FEB-1996.
XX
PF   28-JUL-1995;   95WO-US09599.
XX
PR   29-JUL-1994;   94US-0282758.
XX
PA   (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA   (USSH) US SEC DEPT HEALTH.
XX
PI   Fields HA, Khudyakov YE;
XX
DR   WPI; 1996-129330/13.
XX

PN Antigenic peptide(s) binding anti-hepatitis C virus antibodies -
PT useful for differential diagnosis of HCV in subjects
XX
PS Example; Page 39; 50pp; English.
XX
CC The present peptide, which was derived from the hepatitis C virus
CC (HCV) non-structural protein 4 (NS4), was prepd. using FMOC
CC chemistry. It was tested for reactivity against a panel of 32
CC anti-HCV positive sera, and was found to react with 15 of them.
XX
SQ Sequence 10 AA;

Query Match   100.0%; Score 23; DB 17; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

     1 VPGG 4
       ||||
     7 vpgg 10

RESULT 14
W41471
ID W41471 standard; peptide; 10 AA.
XX
AC W41471;
XX
DT 05-JUN-1998 (first entry)
XX
DE Fragment of nematode alpha 3-tubulin.
XX
KW Nematode; alpha 3-tubulin; aligned peptide array; protein binding site;
KW protein interaction site; ligand detection.
XX
OS Nematoda.
XX
PN EP818467-A2.
XX
PD 14-JAN-1998.
XX
PF 11-JUL-1997; 97EP-0111868.
XX
PR 12-JUL-1996; 96JP-0183140.
XX
PA (NIDE) NBC CORP.
XX
PI Miwa J, Siddiqui SS;
XX
DR WPI; 1998-065262/07.
XX
PT Aligned peptide array - for detecting protein binding or interaction
PT sites, etc.
XX
PS Example 1; Page 5; 27pp; English.
XX
CC This sequence represents a fragment of alpha 3-tubulin derived from a
CC nematode. This sequence was used as a part of an aligned peptide array of
CC the invention. The aligned peptide array comprises, as separate elements,
CC peptide segments obtained by dividing the amino acid sequence of a
CC protein into sequence segments of any suitable length and of any suitable
CC frame and synthesizing peptides on the basis of the sequence segments,
CC preferably with the amino acid sequences of the peptide segments
CC expressing the amino acid sequence of the protein. The peptide array can
CC be used in a method for the detection of a binding or interaction site of
CC a protein; in a method for the detection of a ligand for a protein; in a
CC method for the modification of a protein; and in a method for the design
CC of a protein. Using the information from the above methods, ligands
CC detected by them can be modified or designed. The peptide array can also
CC be used in enzyme immunoassay method, represented by ELISA. The array can
CC be used in a vizoimmunoassay method, a metalloimmunoassay method, a
CC fluoroimmunoassay method or, a radioimmunoassay method.
XX

SQ Sequence 10 AA;

Query Match   100.0%; Score 23; DB 19; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 VPGG 4
         ||||
Db   5 vpgg 8

RESULT 15
Y31331
ID Y31331 standard; peptide; 11 AA.
XX
AC Y31331;
XX
DT 01-NOV-1999 (first entry)
XX
DE B. subtilis surface binding peptide ligand.
XX
KW Bacterial spore; biopanning; phage-display library; Bacillus; vaccine;
KW pathogen detection; biological warfare agent; B. anthracis.
XX
OS Bacillus subtilis.
XX
PN WO9936081-A1.
XX
PD 22-JUL-1999.
XX
PF 14-JAN-1999; 99WO-US00771.
XX
PR 14-JAN-1998; 98US-0071411.
XX
PA (UABR-) UAB RES FOUND.
XX
PI Turnbough CL;
XX
DR WPI; 1999-468943/39.
XX
PT Identifying peptides that bind to the surface of bacterial spores by
PT biopanning phage-display library, useful as vaccines and diagnostic
PT agents
XX
PS Disclosure; Page 9; 23pp; English.
XX
CC The invention provides peptides that bind to the surface of bacterial
CC spores. These peptides are identified by biopanning a phage-display
CC library with the spores. The library is incubated with spores and any
CC phage-spore complexes formed recovered by centrifuging. They are washed
CC thoroughly, then phage eluted with buffer, the eluate neutralised and
CC eluted phage amplified. The procedure is repeated, for 3 or 4 rounds of
CC panning, then individual clones purified, amplified and genomic DNA
CC extracted for determination of peptide-encoding sequences. Peptides
CC encoded by these sequences are then tested for ability to bind to target
CC spores. The peptides are used for capture and identification of bacterial
CC spores, particularly of the genus Bacillus, particularly for detecting
CC pathogens, or otherwise harmful species, in the environment (e.g. air,
CC water or food) or in clinical samples. They are also used to protect
CC against disease-causing spores (e.g. by incorporation in protective
CC masks), and in vaccines to generate a protective antibody response. The
CC peptides-specific antibodies are useful as detection reagents, e.g. in
CC enzyme-linked immunosorbent assay. The peptides are species specific,
CC i.e. they can differentiate between the possible biological warfare agent
CC B. anthracis and related species widely present in the environment and
CC likely to give rise to false positive results. The present sequence
CC represents a peptide that can bind to B. subtilis.
XX
SQ Sequence 11 AA;

Query Match   100.0%; Score 23; DB 20; Length 11;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
          ||||
Db   6 vpgg 9

Search completed: April 24, 2001, 16:38:20
Job time: 420 sec

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        April 24, 2001, 16:42:49 ; Search time 44.88 seconds
                                          (without alignments)
                                          3.816 Million cell updates/sec

Title:         US-09-340-736-6
Perfect score: 27
Sequence:      1 VPGVG 5

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

searched:       93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters:    93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                        SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	27	100.0	93	1	YM03_STRCM	Q05072 streptomyce
2	27	100.0	128	1	COXE_CAEEL	Q20079 caenorhabdi
3	27	100.0	181	1	Y65L_HORVU	O48609 hordeum vul
4	27	100.0	192	1	KADA_METIG	P43408 methanococc
5	27	100.0	192	1	KADA_METJA	P43409 methanococc
6	27	100.0	192	1	KADA_METTL	P43410 methanococc
7	27	100.0	192	1	KADA_METVO	P43411 methanococc
8	27	100.0	204	1	KADA_ASRPE	Q9yd42 aeropyrum p
9	27	100.0	206	1	HISS_MYCLE	Q9x7c0 mycobacteri
10	27	100.0	206	1	HISS_MYCTU	O06589 mycobacteri
11	27	100.0	211	1	HISS_CORGL	O69043 corynebacte
12	27	100.0	222	1	HISS_STRCO	P16249 streptomyce
13	27	100.0	255	1	YX37_MYCTU	Q10519 mycobacteri
14	27	100.0	274	1	DCOP_MYCTU	P77698 mycobacteri
15	27	100.0	276	1	DCOP_MYCSM	O08323 mycobacteri
16	27	100.0	281	1	YN07_MYCTU	Q50658 mycobacteri
17	27	100.0	283	1	PHAR_PSEOL	P26495 pseudomonas
18	27	100.0	284	1	STAR_RAT	P97826 rattus norv
19	27	100.0	294	1	YQXK_BACSU	P38423 bacillus su
20	27	100.0	324	1	CCS1_CAEEL	P12214 caenorhabdi
21	27	100.0	334	1	YE96_MYCTU	P71777 mycobacteri
22	27	100.0	374	1	OMPF_SERMA	O33980 serratia ma
23	27	100.0	393	1	EX53_MYCTU	Q10699 mycobacteri
24	27	100.0	402	1	YIN2_STRAM	P32426 streptomyce
25	27	100.0	421	1	SYH_THETH	P56194 thermus aqu
26	27	100.0	478	1	GSHB_ARATH	P46416 arabidopsis
27	27	100.0	479	1	CATA_PSEPU	Q59714 pseudomonas
28	27	100.0	492	1	CATA_ONCVE	Q27710 onchocerca
29	27	100.0	488	1	U2AF_CAEBR	P90727 caenorhabdi
30	27	100.0	496	1	U2AF_CAEEL	P90978 caenorhabdi
31	27	100.0	510	1	CP46_ORYCU	P14580 oryctolagus
32	27	100.0	511	1	CP47_RABIT	P14581 oryctolagus
33	27	100.0	519	1	CBX2_MOUSE	P30658 mus musculu
34	27	100.0	566	1	CAG3_CHICK	Q92183 gallus gall
35	27	100.0	582	1	HEXA_MUMP1	P19762 mumps virus
36	27	100.0	582	1	HEXA_MUMPM	P11235 mumps virus
37	27	100.0	582	1	HEXA_MUMPR	P10866 mumps virus
38	27	100.0	583	1	GTB1_MOUSE	O08582 mus musculu
39	27	100.0	584	1	GTB1_HUMAN	O00178 homo sapien
40	27	100.0	629	1	YS50_MYCTU	O05809 mycobacteri
41	27	100.0	649	1	GUND_CLOTM	P04954 clostridium
42	27	100.0	730	1	ELS_HUMAN	P15502 homo sapien
43	27	100.0	734	1	VTPR_HSVEB	P28969 equine herp
44	27	100.0	735	1	VTPR_HSV11	P04295 herpes simp
45	27	100.0	747	1	ELS_BOVIN	P04985 bos taurus

                        ALIGNMENTS

```
RESULT_1
YM03_STRCM

ID   YM03_STRCM      STANDARD;      PRT;    93 AA.
AC   Q05072;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   HYPOTHETICAL PROTEIN IN MUTB 3'REGION (ORF-C) (FRAGMENT).
OS   Streptomyces cinnamonensis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1900;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3823.5;
RX   MEDLINE=93273720; PubMed=8090072;
RA   Birch A., Leiser A., Robinson J.A.;
RT   "Cloning, sequencing, and expression of the gene encoding
RT   methylmalonyl-coenzyme A mutase from Streptomyces cinnamonensis.";
RL   J. Bacteriol. 175:3511-3519(1993).
CC   -!- SIMILARITY: BELONGS TO THE ARGK FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L10064; AAA03042.1; -.
DR   PIR; C40595; C40595.
KW   Hypothetical protein.
FT   NON_TER         93     93
SQ   SEQUENCE     93 AA;   9858 MW;  E667DD645B3845D3 CRC64;

Query Match             100.0%;  Score 27;  DB 1;  Length 93;
Best Local Similarity   100.0%;  Pred. No. 51;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VPGVG 5
              |||||
Db         41 VPGVG 45


RESULT_2
COXE_CAEEL

ID   COXE_CAEEL      STANDARD;      PRT;    128 AA.
AC   Q20079;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PROBABLE CYTOCHROME C OXIDASE POLYPEPTIDE VIA PRECURSOR (EC 1.9.3.1).
GN   F54D8.2.
OS   Caenorhabditis elegans.
```

```
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Bentley D.;
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE
CC       CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN
CC       MITOCHONDRIAL ELECTRON TRANSPORT (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VIA FAMILY.
CC       -----------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       -----------------------------------------------------------------
DR   EMBL; U12966; AAA20614.1; -.
DR   WormPep; F54D8.2; CE01508.
DR   InterPro; IPR001349; -.
DR   Pfam; PF02046; COX6A; 1.
DR   PROSITE; PS01329; COX6A; 1.
KW   Oxidoreductase; Inner membrane; Mitochondrion; Transit peptide.
FT   TRANSIT       1       ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?     128       PROBABLE CYTOCHROME C OXIDASE POLYPEPTIDE
FT                                 VIA.
SQ   SEQUENCE   128 AA;  14743 MW;  EF4E4A56A1CB6A233 CRC64;

Query Match             100.0%;  Score 27;  DB 1;  Length 128;
Best Local Similarity   100.0%;  Pred. No. 68;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 VPGVG 5
         |||||
Db   116 VPGVG 120

RESULT 3
ID   Y65L_HORVU     STANDARD;      PRT;     181 AA.
AC   O48609;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   YCP65-LIKE PROTEIN PRECURSOR.
OS   Hordeum vulgare (Barley).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
OC   Hordeum.
OX   NCBI_TaxID=4513;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. HAISA; TISSUE=Leaf;
RA   Hess W.R., Golz R., Boerner T.;
RT   "Analysis of randomly selected cDNAs reveals the expression of stress-
RT   and defence-related genes in the barley mutant albostrians.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE YCF65 FAMILY.
CC       -----------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
```

```
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       -----------------------------------------------------------------
DR   EMBL; AJ222779; CAA10984.1; -.
KW   Hypothetical protein; Chloroplast; Transit peptide.
FT   TRANSIT       1       ?       CHLOROPLAST (POTENTIAL).
FT   CHAIN         ?     181       YCF65-LIKE PROTEIN
SQ   SEQUENCE   181 AA;  19865 MW;  B02DAC3792F73BB5 CRC64;

Query Match             100.0%;  Score 27;  DB 1;  Length 181;
Best Local Similarity   100.0%;  Pred. No. 94;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 VPGVG 5
         |||||
Db   120 VPGVG 124

RESULT 4
ID   KADA_METIG     STANDARD;      PRT;     192 AA.
AC   P43408;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).
GN   ADKA OR ADK.
OS   Methanococcus igneus.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2189;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97208879; PubMed=9055821;
RA   Ferber D.M., Haney P.J., Berk H., Lynn D., Konisky J.;
RT   "The adenylate kinase genes of M. voltae, M. thermolithotrophicus, M.
RT   jannaschii, and M. igneus define a new family of adenylate kinases.";
RL   Gene 185:239-244(1997).
RN   [2]
RP   SEQUENCE OF 1-30.
RX   MEDLINE=95286473; PubMed=7768791;
RA   Rusnak P., Haney P., Konisky J.;
RT   "The adenylate kinases from a mesophilic and three thermophilic
RT   methanogenic members of the Archaea.";
RL   J. Bacteriol. 177:2977-2981(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC   -!- SUBUNIT: MONOMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- MISCELLANEOUS: ACTIVE IN THE TEMPERATURE RANGE OF 70 TO 90 DEGREES
CC       CELSIUS.
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEAL ADENYLATE KINASE FAMILY.
CC       -----------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       -----------------------------------------------------------------
DR   EMBL; U39881; AAC44862.1; -.
KW   Transferase; Kinase; ATP-binding.
FT   NP_BIND      10      18       ATP (BY SIMILARITY).
FT   CONFLICT      8       8       V -> I (IN REF. 2).
FT   CONFLICT     20      20       T -> L (IN REF. 2).
SQ   SEQUENCE   192 AA;  21400 MW;  AB2EC9C9DB905E75 CRC64;

Query Match             100.0%;  Score 27;  DB 1;  Length 192;
Best Local Similarity   100.0%;  Pred. No. 1e+02;
Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

Qy   1 VPGVG 5
          |||||
Db   11 VPGVG 15

RESULT 5
KADA_METJA
ID   KADA_METJA   STANDARD;   PRT;   192 AA.
AC   P43409;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).
GN   ADKA OR ADK OR MJ0479.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97208879; PubMed=9055821;
RA   Ferber D.M., Haney P.J., Berk H., Lynn D., Konisky J.;
RT   "The adenylate kinase genes of M. voltae, M. thermolithotrophicus, M.
RT   jannaschii, and M. igneus define a new family of adenylate kinases.";
RL   Gene 185:239-244(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olson G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
RN   [3]
RP   SEQUENCE OF 1-30.
RX   MEDLINE=95286473; PubMed=7768791;
RA   Rusnak P., Haney P., Konisky J.;
RT   "The adenylate kinases from a mesophilic and three thermophilic
RT   methanogenic members of the Archaea.";
RL   J. Bacteriol. 177:2977-2981(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC   -!- SUBUNIT: MONOMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- MISCELLANEOUS: ACTIVE IN THE TEMPERATURE RANGE OF 70 TO 90 DEGREES
CC   CELSIUS.
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEAL ADENYLATE KINASE FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U39882; AAC44863.1; -.
DR   EMBL; U67498; AAB98470.1; ALT_INIT.
DR   TIGR; MJ0479; -.
KW   Transferase; Kinase; ATP-binding.
FT   NP_BIND   10   18   ATP (BY SIMILARITY).
FT   CONFLICT   15   15   S -> G (IN REF. 3).
SQ   SEQUENCE   192 AA; 21772 MW; ECD533AD4C85D99BE CRC64;

Query Match   100.0%; Score 27; DB 1; Length 192;

Best Local Similarity   100.0%; Pred. No. 1e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   11 VPGVG 15

RESULT 6
KADA_METTL
ID   KADA_METTL   STANDARD;   PRT;   192 AA.
AC   P43410;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).
GN   ADKA OR ADK.
OS   Methanococcus thermolithotrophicus.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2186;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97208879; PubMed=9055821;
RA   Ferber D.M., Haney P.J., Berk H., Lynn D., Konisky J.;
RT   "The adenylate kinase genes of M. voltae, M. thermolithotrophicus, M.
RT   jannaschii, and M. igneus define a new family of adenylate kinases.";
RL   Gene 185:239-244(1997).
RN   [2]
RP   SEQUENCE OF 1-20.
RX   MEDLINE=95286473; PubMed=7768791;
RA   Rusnak P., Haney P., Konisky J.;
RT   "The adenylate kinases from a mesophilic and three thermophilic
RT   methanogenic members of the Archaea.";
RL   J. Bacteriol. 177:2977-2981(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC   -!- SUBUNIT: MONOMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- MISCELLANEOUS: ACTIVE IN THE TEMPERATURE RANGE OF 60 TO 80 DEGREES
CC   CELSIUS.
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEAL ADENYLATE KINASE FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U39880; AAC44864.1; -.
KW   Transferase; Kinase; ATP-binding.
FT   NP_BIND   10   18   ATP (BY SIMILARITY).
SQ   SEQUENCE   192 AA; 21461 MW; 7223378B43320B1 CRC64;

Query Match   100.0%; Score 27; DB 1; Length 192;
Best Local Similarity   100.0%; Pred. No. 1e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   11 VPGVG 15

RESULT 7
KADA_METVO
ID   KADA_METVO   STANDARD;   PRT;   192 AA.
AC   P43411;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).

GN   ADKA OR ADK.
OS   Methanococcus voltae.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2188;
RN   [1]
RC   STRAIN=PS;
RX   MEDLINE=97208879; PubMed=9055821;
RA   Ferber D.M., Haney P.J., Berk H., Lynn D., Konisky J.;
RT   "The adenylate kinase genes of M. voltae, M. thermolithotrophicus, M.
RT   jannaschii and M igneus define a new family of adenylate kinases.";
RL   Gene 185:239-244(1997).
RN   [2]
RX   MEDLINE=95286473; PubMed=7768791;
RA   Rusnak P., Haney P., Konisky J.;
RT   "The adenylate kinases from a mesophilic and three thermophilic
RT   methanogenic members of the Archaea.";
RL   J. Bacteriol. 177:2977-2981(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC   -!- SUBUNIT: MONOMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- MISCELLANEOUS: ACTIVE IN THE TEMPERATURE RANGE OF 30 TO 40 DEGREES
CC   CELSIUS.
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEAL ADENYLATE KINASE FAMILY.
CC   -!-
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U39879; AAC44865.1; -.
KW   Transferase; Kinase; ATP-binding.
FT   NP_BIND   10   18   ATP (BY SIMILARITY).
FT   CONFLICT   31   31   G -> GG (IN REF. 2).
SQ   SEQUENCE   192 AA; 21303 MW; 952ABCD178BD6A8E CRC64;

Query Match   100.0%; Score 27; DB 1; Length 192;
Best Local Similarity   100.0%; Pred. No. 1e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   11 VPGVG 15

RESULT 8
KADA_AERPE
ID   KADA_AERPE   STANDARD;   PRT;   204 AA.
AC   Q9YDD2;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).
GN   ADKA OR APE0981.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A, Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1 ";
RL   DNA Res. 6:83-101(1999).
CC   -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEAL ADENYLATE KINASE FAMILY.
CC   -!-
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AP000060; BAA79965.1; -.
KW   Transferase; Kinase; ATP-binding.
FT   NP_BIND   14   22   ATP (BY SIMILARITY).
SQ   SEQUENCE   204 AA; 22175 MW; 279EE5F8BD381D13 CRC64;

Query Match   100.0%; Score 27; DB 1; Length 204;
Best Local Similarity   100.0%; Pred. No. 1.1e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   15 VPGVG 19

RESULT 9
HIS5_MYCLE
ID   HIS5_MYCLE   STANDARD;   PRT;   206 AA.
AC   Q9X7C0;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   AMIDOTRANSFERASE HISH (EC 2.4.2.-).
GN   HISH OR MLCB1610.23.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Seeger K.J., Harris D., James K.D., Parkhill J., Barrell B.G.,
RA   Rajandream M.A.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYZES AN AMIDOTRANSFERASE REACTION THAT GENERATES
CC   IMIDAZOLE-GLYCEROL PHOSPHATE AND 5-AMINOIMIDAZOL-4-CARBOXAMIDE
CC   RIBONUCLEOTIDE, WHICH IS USED FOR PURINE SYNTHESIS.
CC   -!- PATHWAY: FIFTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE HISH FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC   -!-
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AL049913; CAB43169.1; -.
DR   InterPro; IPR000991; -.
DR   Pfam; PF00117; GATase; 1.
DR   PROSITE; PS00442; GATASE_TYPE_I; 1.
KW   Histidine biosynthesis; Transferase; Glutamine amidotransferase.
FT   ACT_SITE   83   83   BY SIMILARITY.
FT   ACT_SITE   187   187   BY SIMILARITY.
FT   ACT_SITE   189   189   BY SIMILARITY.

SQ   SEQUENCE   206 AA;   21652 MW; 72D699408AF81536 CRC64;

Query Match   100.0%;   Score 27; DB 1; Length 206;
Best Local Similarity   100.0%;   Pred. No. 1.1e+02;
Matches   5;   Conservative   0;   Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   47 VPGVG 51

RESULT 10
HIS5_MYCTU
ID   HIS5_MYCTU   STANDARD;   PRT;   206 AA.
AC   O06589;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   AMIDOTRANSFERASE HISH (EC 2.4.2.-).
GN   HISH OR RV1602 OR MTCY336.02C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares R., Squares S., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: CATALYZES AN AMIDOTRANSFERASE REACTION THAT GENERATES
CC   IMIDAZOLE-GLYCEROL PHOSPHATE AND 5-AMINOIMIDAZOL-4-CARBOXAMIDE
CC   RIBONUCLEOTIDE, WHICH IS USED FOR PURINE SYNTHESIS.
CC   -!- PATHWAY: FIFTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE HISH FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Z95586; CAB09092.1; -.
DR   TubercuList; Rv1602; -.
DR   InterPro; IPR000991; -.
DR   Pfam; PF00117; GATase; 1.
DR   PROSITE; PS00442; GATASE_TYPE_I; 1.
KW   Histidine biosynthesis; Transferase; Glutamine amidotransferase.
FT   ACT_SITE   83   83   BY SIMILARITY.
FT   ACT_SITE   187   187   BY SIMILARITY.
FT   ACT_SITE   189   189   BY SIMILARITY.
SQ   SEQUENCE   206 AA;   21418 MW; 4B2AF9C61BEE1447 CRC64;

Query Match   100.0%;   Score 27; DB 1; Length 206;
Best Local Similarity   100.0%;   Pred. No. 1.1e+02;
Matches   5;   Conservative   0;   Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   47 VPGVG 51

RESULT 11
HIS5_CORGL
ID   HIS5_CORGL   STANDARD;   PRT;   211 AA.
AC   O69043;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   AMIDOTRANSFERASE HISH (EC 2.4.2.-).
GN   HISH.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC   Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13059 / AS019;
RA   Juns S.I., Han M.S., Park Y.J., Lee S.K., Lee M.S.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYZES AN AMIDOTRANSFERASE REACTION THAT GENERATES
CC   IMIDAZOLE-GLYCEROL PHOSPHATE AND 5-AMINOIMIDAZOL-4-CARBOXAMIDE
CC   RIBONUCLEOTIDE, WHICH IS USED FOR PURINE SYNTHESIS.
CC   -!- PATHWAY: FIFTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE HISH FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AF060558; AAC15231.1; -.
DR   InterPro; IPR000991; -.
DR   Pfam; PF00117; GATase; 1.
DR   PROSITE; PS00442; GATASE_TYPE_I; 1.
KW   Histidine biosynthesis; Transferase; Glutamine amidotransferase.
FT   ACT_SITE   82   82   BY SIMILARITY.
FT   ACT_SITE   190   190   BY SIMILARITY.
FT   ACT_SITE   192   192   BY SIMILARITY.
SQ   SEQUENCE   211 AA;   23139 MW; A935FEC422B48AB1 CRC64;

Query Match   100.0%;   Score 27; DB 1; Length 211;
Best Local Similarity   100.0%;   Pred. No. 1.1e+02;
Matches   5;   Conservative   0;   Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
          |||||
Db   47 VPGVG 51

RESULT 12
HIS5_STRCO
ID   HIS5_STRCO   STANDARD;   PRT;   222 AA.
AC   P16249;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   AMIDOTRANSFERASE HISH (EC 2.4.2.-).
GN   HISH OR SC4G6.20C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]

RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=90337345; PubMed=2199329;
RA   Limauro D., Avitabile A., Cappellano M., Puglia A.M., Bruni C.B.;
RT   "Cloning and characterization of the histidine biosynthetic gene
RT   cluster of Streptomyces coelicolor A3(2).";
RL   Gene 90:31-41(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D., Bentley S.D., Parkhill J., Barrell B.G.,
RA   Rajandream M.A.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CATALYZES AN AMIDOTRANSFERASE REACTION THAT GENERATES
CC   IMIDAZOLE-GLYCEROL PHOSPHATE AND 5-AMINOIMIDAZOL-4-CARBOXAMIDE
CC   RIBONUCLEOTIDE, WHICH IS USED FOR PURINE SYNTHESIS.
CC   -!- PATHWAY: FIFTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE HISH FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC   --
CC   --
DR   EMBL; M31528; AAA26759.1; -.
DR   EMBL; AL096884; CAB51443.1; -.
DR   PIR; JQ0640; JQ0640.
DR   InterPro; IPR000991; -.
DR   Pfam; PF00117; GATase; 1.
DR   PROSITE; PS00442; GATASE_TYPE_1; 1.
KW   Histidine biosynthesis; Transferase; Glutamine amidotransferase.
FT   ACT_SITE   93   93   BY SIMILARITY.
FT   ACT_SITE   203   203   BY SIMILARITY.
FT   ACT_SITE   205   205   BY SIMILARITY.
SQ   SEQUENCE   222 AA; 23861 MW; C075C843547A4CB6 CRC64;

Query Match   100.0%; Score 27; DB 1; Pred. No. 1.1e+02; Length 222;
Best Local Similarity   100.0%;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   57 VPGVG 61

RESULT 13
YM37_MYCTU
ID   YM37_MYCTU   STANDARD;   PRT;   255 AA.
AC   Q10519;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL 29.1 KDA PROTEIN RV2237.
GN   RV2237 OR MTCY427.18.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   --
CC   --
DR   EMBL; Z70692; CAA94667.1; -.
DR   Tuberculist; RV2237; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM   2   22   POTENTIAL.
FT   TRANSMEM   168   188   POTENTIAL.
SQ   SEQUENCE   255 AA; 29067 MW; B0F422277C6B46A6 CRC64;

Query Match   100.0%; Score 27; DB 1; Length 255;
Best Local Similarity   100.0%; Pred. No. 1.3e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGVG 5
              |||||
Db   14 VPGVG 18

RESULT 14
DCOP_MYCTU
ID   DCOP_MYCTU   STANDARD;   PRT;   274 AA.
AC   P77898; P42610;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   OROTIDINE 5'-PHOSPHATE DECARBOXYLASE (EC 4.1.1.23) (OMP
DE   DECARBOXYLASE) (OXMPDCASE).
GN   PYRF OR URAA OR RV1385 OR MTCY1B4.02.
OS   Mycobacterium tuberculosis, and Mycobacterium bovis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773, 1765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.tuberculosis; STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=M.bovis; STRAIN=BCG;
RX   MEDLINE=94042902; PubMed=8226675;
RA   Aldovini A., Husson R.N., Young R.A.;
RT   "The uraA locus and homologous recombination in Mycobacterium bovis
RT   BCG.";
RL   J. Bacteriol. 175:7282-7289(1993).
CC   -!- CATALYTIC ACTIVITY: OROTIDINE-5'-PHOSPHATE = UMP + CO(2).
CC   -!- PATHWAY: SIXTH AND LAST STEP IN THE BIOSYNTHESIS OF PYRIMIDINES.
CC   -!- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY.
CC   --

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z80108; CAB02190.1; -.
DR   EMBL; U01072; AAC43182.1; -.
DR   TuberculList; Rv1385.
DR   InterPro; IPR001754; -.
DR   Pfam; PF00215; OMPdecase; 1.
DR   PROSITE; PS00156; OMPDECASE; 1.
KW   Pyrimidine biosynthesis; Lyase; Decarboxylase.
FT   ACT_SITE       95     95       BY SIMILARITY.
SQ   SEQUENCE    274 AA;  27377 MW;  3698AE076FB3D143 CRC64;

Query Match             100.0%;  Score 27;  DB 1;  Length 274;
Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
Matches    5; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

Qy          1 VPGVG 5
              |||||
Db        215 VPGVG 219
```

```
RESULT 15
DCOP_MYCSM
ID   DCOP_MYCSM     STANDARD;      PRT;    276 AA.
AC   O08323;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   OROTIDINE 5'-PHOSPHATE DECARBOXYLASE (EC 4.1.1.23) (OMP
DE   DECARBOXYLASE) (OMPDCASE).
GN   PYRF.
OS   Mycobacterium smegmatis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1772;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 700084 / MC(2)155;
RX   MEDLINE=97312763; PubMed=9169204;
RA   Knipfer N., Seth A., Shrader T.E.;
RT   "Unmarked gene integration into the chromosome of Mycobacterium
RT   smegmatis via precise replacement of the pyrF gene.";
RL   Plasmid 37:129-140(1997).
CC   -!- CATALYTIC ACTIVITY: OROTIDINE-5'-PHOSPHATE = UMP + CO(2).
CC   -!- PATHWAY: SIXTH AND LAST STEP IN THE BIOSYNTHESIS OF PYRIMIDINES.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U91572; AAB50157.1; -.
DR   InterPro; IPR001754; -.
DR   Pfam; PF00215; OMPdecase; 1.
DR   PROSITE; PS00156; OMPDECASE; 1.
KW   Pyrimidine biosynthesis; Lyase; Decarboxylase.
FT   ACT_SITE       95     95       BY SIMILARITY.
SQ   SEQUENCE    276 AA;  27815 MW;  41437843A352A896 CRC64;

Query Match             100.0%;  Score 27;  DB 1;  Length 276;
```

Best Local Similarity   100.0%; Pred. No. 1.4e+02;
Matches   5; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          1 VPGVG 5
              |||||
Db        221 VPGVG 225
```

Search completed: April 24, 2001, 16:42:50
Job time: 456 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:36:24 ; Search time 62.39 Seconds
                                        (without alignments)
                                        1.232 Million cell updates/sec

Title: US-09-340-736-7
Perfect score: 23
Sequence: 1 VPGG 4

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_AA:*
                 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	23	100.0	23	4	1	US-08-106-509-7	Sequence 7, Appli
2	23	100.0	23	4	1	US-07-609-716-1	Sequence 1, Appli
3	23	100.0	23	4	2	US-08-483-236-18	Sequence 18, Appl
4	23	100.0	23	4	2	US-08-911-364-7	Sequence 7, Appli
5	23	100.0	23	4	2	US-08-735-692-9	Sequence 9, Appli
6	23	100.0	23	4	4	US-08-542-051-3	Sequence 3, Appli
7	23	100.0	23	4	4	US-08-963-168C-29	Sequence 29, Appli
8	23	100.0	23	4	4	US-08-475-411A-1	Sequence 1, Appli
9	23	100.0	23	4	4	US-08-478-029A-1	Sequence 1, Appli
10	23	100.0	23	6	2	5250516-3	Patent No. 5250516
11	23	100.0	23	7	4	US-08-340-283-182	Sequence 182, App
12	23	100.0	23	7	4	US-09-147-933-29	Sequence 29, Appl
13	23	100.0	23	9	3	US-08-340-283-64	Sequence 64, Appl
14	23	100.0	23	10	1	US-08-282-758B-32	Sequence 32, Appl
15	23	100.0	23	12	1	US-08-282-758B-14	Sequence 14, Appl
16	23	100.0	23	13	1	US-08-290-448A-11	Sequence 11, Appli
17	23	100.0	23	13	1	US-08-290-448A-11	Sequence 11, Appli
18	23	100.0	23	13	1	US-08-175-069A-11	Sequence 7, Appli
19	23	100.0	23	14	2	US-08-258-951-7	Sequence 76, Appli
20	23	100.0	23	14	2	US-08-529-190B-76	Sequence 76, Appli
21	23	100.0	23	16	1	US-07-789-912-5	Sequence 5, Appli
22	23	100.0	23	20	1	US-08-837-226-12	Sequence 12, Appli
23	23	100.0	23	27	1	US-07-938-782A-11	Sequence 11, Appli
24	23	100.0	23	27	1	US-08-630-524-11	Sequence 11, Appl
25	23	100.0	23	43	2	US-08-488-161-78	Sequence 78, Appli
26	23	100.0	23	43	3	US-09-273-685-78	Sequence 78, Appli
27	23	100.0	23	43	5	PCT-US95-11934-78	Sequence 78, Appli
28	23	100.0	53	1	US-08-765-957-1	Sequence 1, Appli
29	23	100.0	54	1	US-08-485-455D-39	Sequence 39, Appl
30	23	100.0	54	2	US-08-482-130C-39	Sequence 39, Appl
31	23	100.0	54	2	US-08-484-211C-39	Sequence 39, Appl
32	23	100.0	54	3	US-08-906-769-39	Sequence 39, Appl
33	23	100.0	54	4	US-08-906-616-39	Sequence 39, Appl
34	23	100.0	54	4	US-08-817-795-39	Sequence 39, Appl
35	23	100.0	54	4	US-08-485-443B-39	Sequence 39, Appl
36	23	100.0	54	4	US-08-639-075A-39	Sequence 39, Appl
37	23	100.0	54	4	US-09-012-431-39	Sequence 39, Appl
38	23	100.0	54	5	PCT-US95-14442A-39	Sequence 39, Appl
39	23	100.0	57	4	US-08-506-553C-1	Sequence 1, Appli
40	23	100.0	77	4	US-09-025-151-20	Sequence 20, Appl
41	23	100.0	84	3	US-08-906-769-99	Sequence 99, Appl
42	23	100.0	84	4	US-08-906-616-99	Sequence 99, Appl
43	23	100.0	84	4	US-08-817-795-99	Sequence 99, Appl
44	23	100.0	84	4	US-08-639-075A-99	Sequence 99, Appl
45	23	100.0	84	4	US-09-012-431-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1
US-08-106-509-7
; Sequence 7, Application US/08106509
; Patent No. 5428014
; GENERAL INFORMATION:
; APPLICANT: Labroo, Virender
; APPLICANT: Busby, Sharon J.
; TITLE OF INVENTION: TRANSGLUTAMINASE CROSS-LINKABLE
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS RELATING THERETO
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: ZymoGenetics, Inc.
;   STREET: 4225 Roosevelt Way, N.E.
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98105
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/106,509
;   FILING DATE: 19930813
;   CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
;   NAME: Parker, Gary E
;   REGISTRATION NUMBER: 31-684
;   REFERENCE/DOCKET NUMBER: 93-09
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-547-8080 ext 322
;   TELEFAX: 206-548-2329
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 4 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   FRAGMENT TYPE: internal
US-08-106-509-7

Query Match   100.0%; Score 23; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
            ||||
Db   1 VPGG 4

RESULT 2
US-07-609-716-1
; Sequence 1, Application US/07609716
; Patent No. 551581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-609-716-1

Query Match   100.0%; Score 23; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
              ||||
Db   1 VPGG 4

RESULT 3
US-08-483-236-18
; Sequence 18, Application US/08483236
; Patent No. 5939385
; GENERAL INFORMATION:
; APPLICANT: Labroo, Virender
; APPLICANT: Busby, Sharon
; TITLE OF INVENTION: Transglutaminase Cross-linkable
; TITLE OF INVENTION: Polypeptides and Methods Relating Thereto
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,236
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 93-09cl
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-483-236-18

Query Match   100.0%; Score 23; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
              ||||
Db   1 VPGG 4

RESULT 4
US-08-911-364-7
; Sequence 7, Application US/08911364
; Patent No. 5969106
; GENERAL INFORMATION:
; APPLICANT: ROTHSTEIN, Aser
; APPLICANT: KEELY, Fred W.
; APPLICANT: ROTHSTEIN, Steven J.
; TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
; TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,364
; FILING DATE: 07-AUG-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/023,552
; FILING DATE: 07-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 041082/0104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids

```
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-911-364-7

Query Match        100.0%;  Score 23;  DB 2;  Length 4;
Best Local Similarity  100.0%;  Pred. No. 1.4e+05;
Matches   4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 VPGG 4
        ||||
Db    1 VPGG 4

RESULT 5
US-08-735-692-9
; Sequence 9, Application US/08735692B
; Patent No. 5972406
; GENERAL INFORMATION:
; APPLICANT: Urry, Dan W.
; APPLICANT: Shewry, Peter R.
; APPLICANT: Prasad, Kari
; TITLE OF INVENTION: Bioelastomers Suitable as Food Product Additives
; FILE REFERENCE: BERL-017/01US
; CURRENT APPLICATION NUMBER: US/08/735,692B
; CURRENT FILING DATE: 1995-10-16
; EARLIER APPLICATION NUMBER: 08/423,517
; EARLIER FILING DATE: 1995-04-14
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic
US-08-735-692-9

Query Match        100.0%;  Score 23;  DB 2;  Length 4;
Best Local Similarity  100.0%;  Pred. No. 1.4e+05;
Matches   4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 VPGG 4
        ||||
Db    1 VPGG 4

RESULT 6
US-08-542-051-3
; Sequence 3, Application US/08542051F
; Patent No. 6004782
; GENERAL INFORMATION:
; APPLICANT: Daniell, Henry
; APPLICANT: McPherson, David T.
; APPLICANT: Urry, Dan W.
; APPLICANT: Xu, Jie
; TITLE OF INVENTION: Hyperexpression of Bioelastic Polypeptides
; FILE REFERENCE: BERL-018/01US
; CURRENT APPLICATION NUMBER: US/08/542,051F
; CURRENT FILING DATE: 1995-10-13
; EARLIER APPLICATION NUMBER: 08/423,642
; EARLIER FILING DATE: 1995-04-14
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic
```

```
US-08-542-051-3

Query Match        100.0%;  Score 23;  DB 3;  Length 4;
Best Local Similarity  100.0%;  Pred. No. 1.4e+05;
Matches   4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 VPGG 4
        ||||
Db    1 VPGG 4

RESULT 7
US-08-963-168C-29
; Sequence 29, Application US/08963168C
; Patent No. 6127166
; GENERAL INFORMATION:
; APPLICANT: Bayley, Hagan
; APPLICANT: Cao, Quiping
; APPLICANT: Wang, Yunjaun
; TITLE OF INVENTION: MOLLUSCAN LIGAMENT POLYPEPTIDES
; TITLE OF INVENTION: AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,168C
; FILING DATE: 03-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Passe, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/059001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-963-168C-29

Query Match        100.0%;  Score 23;  DB 4;  Length 4;
Best Local Similarity  100.0%;  Pred. No. 1.4e+05;
Matches   4;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 VPGG 4
        ||||
Db    1 VPGG 4

RESULT 8
US-08-475-411A-1
; Sequence 1, Application US/08475411A
; Patent No. 6140072
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
```

NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
  STREET: Four Embarcadero Center, Suite 3400
  CITY: San Francisco
  STATE: CA
  COUNTRY: US
  ZIP: 94111
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/475,411A
  FILING DATE: 07-JUN-1995
  CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/609,716
  FILING DATE: 06-NOV-1990
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/269,429
  FILING DATE: 09-NOV-1988
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/114,618
  FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 06/927,258
  FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
  NAME: Trecartin, Richard P.
  REGISTRATION NUMBER: 31,801
  REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 415-781-1989
  TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
  LENGTH: 4 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: peptide
US-08-475-411A-1

Query Match   100.0%;   Score 23;   DB 4;   Length 4;
Best Local Similarity 100.0%;   Pred. No. 1.4e+05;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
          ||||
Db   1 VPGG 4

RESULT 9
US-08-478-029A-1
; Sequence 1, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/478,029A
  FILING DATE: 07-JUN-1995
  CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/609,716
  FILING DATE: 06-NOV-1990
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/269,429
  FILING DATE: 09-NOV-1988
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/114,618
  FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 06/927,258
  FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
  NAME: Trecartin, Richard F.
  REGISTRATION NUMBER: 31,801
  REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 415-781-1989
  TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
  LENGTH: 4 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: peptide
US-08-478-029A-1

Query Match   100.0%;   Score 23;   DB 4;   Length 4;
Best Local Similarity 100.0%;   Pred. No. 1.4e+05;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
          ||||
Db   1 VPGG 4

RESULT 10
5250516-3
; Patent No. 5250516
; APPLICANT: URRY, DAN W.
; TITLE OF INVENTION: BIOLASTOMERIC MATERIALS SUITABLE FOR
; THE PROTECTION OF BURN AREAS OR THE PROTECTION OF WOUND
; REPAIR SITES FROM THE OCCURANCE OF ADHESIONS
; NUMBER OF SEQUENCES: 18
; CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/07/184,407
  FILING DATE: 21-APR-1988
; PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 900,895
  FILING DATE: 27-AUG-1986
  APPLICATION NUMBER: 853,212
  FILING DATE: 17-APR-1986
; SEQ ID NO:3:
  LENGTH: 4
5250516-3

Query Match   100.0%;   Score 23;   DB 6;   Length 4;
Best Local Similarity 100.0%;   Pred. No. 1.4e+05;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
          ||||

Db   1 VPGG 4

RESULT 11
US-08-340-283-182
; Sequence 182, Application US/08340283
; Patent No. 5861318
; GENERAL INFORMATION:
;   APPLICANT: Elhammer, Ake P.
;   TITLE OF INVENTION: A SCINTILLATION PROXIMITY ASSAY FOR
;   TITLE OF INVENTION: N-ACETYLGALACTOSAMINYLTRANSFERASE ACTIVITY
;   NUMBER OF SEQUENCES: 205
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Pharmacia and Upjohn, Inc., Intellect. Prop. Law
;   ADDRESSEE: (1920-32-1)
;   STREET: 301 Henrietta Street
;   CITY: Kalamazoo
;   STATE: Michigan
;   COUNTRY: U.S.A.
;   ZIP: 49001
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/340,283
;   FILING DATE:
;   CLASSIFICATION: 436
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Wootton, Thomas A.
;   REGISTRATION NUMBER: 35,004
;   REFERENCE/DOCKET NUMBER: 4828
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (616) 385-7914
;   TELEFAX: (616) 385-6897
;   TELEX: 224401
; INFORMATION FOR SEQ ID NO: 182:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 7 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
;   MOLECULE TYPE: peptide
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FRAGMENT TYPE: N-terminal
US-08-340-283-182

Query Match   100.0%; Score 23; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
        |||
Db   4 VPGG 7

RESULT 14
US-09-147-933-29
; Sequence 29, Application US/09147933A
; Patent No. 6168917
; GENERAL INFORMATION:
;   APPLICANT: Kilpatrick, David
;   TITLE OF INVENTION: DETECTION AND IDENTIFICATION OF
;   TITLE OF INVENTION: NON-POLIO ENTEROVIRUSES
;   FILE REFERENCE: 62242/US
;   CURRENT APPLICATION NUMBER: US/09/147,933A
;   CURRENT FILING DATE: 1999-07-09
;   EARLIER APPLICATION NUMBER: PCT/US97/17734
;   EARLIER FILING DATE: 1997-10-01
;   EARLIER APPLICATION NUMBER: U. S. 60/027,353
; EARLIER FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: peptide
US-09-147-933-29

Query Match   100.0%; Score 23; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
        |||
Db   3 VPGG 6

RESULT 13
US-08-340-283-64
; Sequence 64, Application US/08340283
; Patent No. 5861318
; GENERAL INFORMATION:
;   APPLICANT: Elhammer, Ake P.
;   TITLE OF INVENTION: A SCINTILLATION PROXIMITY ASSAY FOR
;   TITLE OF INVENTION: N-ACETYLGALACTOSAMINYLTRANSFERASE ACTIVITY
;   NUMBER OF SEQUENCES: 205
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Pharmacia and Upjohn, Inc., Intellect. Prop. Law
;   ADDRESSEE: (1920-32-1)
;   STREET: 301 Henrietta Street
;   CITY: Kalamazoo
;   STATE: Michigan
;   COUNTRY: U.S.A.
;   ZIP: 49001
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/340,283
;   FILING DATE:
;   CLASSIFICATION: 436
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Wootton, Thomas A.
;   REGISTRATION NUMBER: 35,004
;   REFERENCE/DOCKET NUMBER: 4828
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (616) 385-7914
;   TELEFAX: (616) 385-6897
;   TELEX: 224401
; INFORMATION FOR SEQ ID NO: 64:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 9 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
;   MOLECULE TYPE: peptide
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FRAGMENT TYPE: N-terminal
US-08-340-283-64

Query Match   100.0%; Score 23; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.4e+05;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4

Db   1 VPGG 4
      ||||

RESULT 14
US-08-282-758B-32
; Sequence 32, Application US/08282758B
; Patent No. 5670310
; GENERAL INFORMATION:
;   APPLICANT: Fields, Howard A.
;   APPLICANT: Khudyakov, Yury E.
;   TITLE OF INVENTION: Methods and Compositions for
;   TITLE OF INVENTION: Differential Diagnosis of Chronic and Acute Hepatitis C
;   TITLE OF INVENTION: Virus Infection
;   NUMBER OF SEQUENCES: 40
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Jones & Askew
;   STREET: 191 Peachtree Street, 37th Floor
;   CITY: Atlanta
;   STATE: Georgia
;   COUNTRY: U.S.A.
;   ZIP: 30303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/282,758B
;   FILING DATE: 29-JUL-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Green, Jamie L.
;   REGISTRATION NUMBER: 32,467
;   REFERENCE/DOCKET NUMBER: 03063-0100
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (404) 818-3700
;   TELEFAX: (404) 818-3799
; INFORMATION FOR SEQ ID NO: 32:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 10 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
US-08-282-758B-32

Query Match   100.0%;   Score 23; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 70;
Matches   4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 VPGG 4
        ||||
Db   7 VPGG 10

RESULT 15
US-08-282-758B-14
; Sequence 14, Application US/08282758B
; Patent No. 5670310
; GENERAL INFORMATION:
;   APPLICANT: Fields, Howard A.
;   APPLICANT: Khudyakov, Yury E.
;   TITLE OF INVENTION: Methods and Compositions for
;   TITLE OF INVENTION: Differential Diagnosis of Chronic and Acute Hepatitis C
;   TITLE OF INVENTION: Virus Infection
;   NUMBER OF SEQUENCES: 40
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Jones & Askew
;   STREET: 191 Peachtree Street, 37th Floor

;   CITY: Atlanta
;   STATE: Georgia
;   COUNTRY: U.S.A.
;   ZIP: 30303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/282,758B
;   FILING DATE: 29-JUL-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Green, Jamie L.
;   REGISTRATION NUMBER: 32,467
;   REFERENCE/DOCKET NUMBER: 03063-0100
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (404) 818-3700
;   TELEFAX: (404) 818-3799
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 12 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
US-08-282-758B-14

Query Match   100.0%;   Score 23; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 83;
Matches   4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 VPGG 4
        ||||
Db   9 VPGG 12

Search completed: April 24, 2001, 16:36:25
Job time: 305 sec

```
              GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

CM protein - protein search, using sw model

Run on:       April 24, 2001, 16:42:03 ; Search time 74.56 Seconds
                                          (without alignments)
                                          3.687 Million cell updates/sec

Title:         US-09-340-736-7
Perfect score: 23
Sequence:      1 VPGG 4

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters:   198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR_67:*
             1: pir1:*
             2: pir2:*
             3: pir3:*
             4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	23	100.0	7	2	B61491	seed protein ws-5
2	23	100.0	15	2	PC2215	fibrinogenolytic p
3	23	100.0	27	2	A28391	hypothetical prote
4	23	100.0	30	2	S14062	hypothetical prote
5	23	100.0	31	2	I52232	tau protein - huma
6	23	100.0	41	2	T36863	probable two-compo
7	23	100.0	45	2	S28397	T-complex protein
8	23	100.0	47	2	A61384	tracheal mucin gly
9	23	100.0	56	2	T36928	hypothetical prote
10	23	100.0	61	2	S03097	aerC protein - Aer
11	23	100.0	62	2	S37135	class II histocomp
12	23	100.0	63	2	T04493	hypothetical prote
13	23	100.0	69	2	F494100	t-complex polypept
14	23	100.0	70	2	T36407	hypothetical prote
15	23	100.0	81	2	B69958	hypothetical prote
16	23	100.0	83	2	S04884	seed protein (clon
17	23	100.0	84	2	T07939	embryonic abundant
18	23	100.0	91	2	S16249	embryonic abundant
19	23	100.0	91	2	T09285	hypothetical prote
20	23	100.0	91	2	T09293	hypothetical prote
21	23	100.0	91	2	T21508	hypothetical prote
22	23	100.0	91	2	D75561	MutT/nudix family
23	23	100.0	92	2	S33790	ribosomal protein
24	23	100.0	92	2	S23527	embryonic abundant
25	23	100.0	92	2	JQ2273	embryonic abundant
26	23	100.0	92	2	S34802	embryonic abunda
27	23	100.0	92	2	S71372	embryogenic abunda
28	23	100.0	93	2	S43333	embryonic abundant
29	23	100.0	93	2	S46303	gB19.1b protein -
30	23	100.0	93	2	S43332	Emb5 protein - whe
31	23	100.0	93	2	S23749	embryonic abundant
32	23	100.0	93	2	A27519	Em protein - wheat
33	23	100.0	93	2	D69416	LSU ribosomal prot
34	23	100.0	93	2	S50649	hypothetical prote
35	23	100.0	93	2	T20227	hypothetical prote
36	23	100.0	93	2	A72653	hypothetical prote
37	23	100.0	94	1	VDWTEM	embryonic abundant
38	23	100.0	95	2	S22483	embryonic abundant
39	23	100.0	96	2	A75308	conserved hypothet
40	23	100.0	97	2	T01984	late-embryogenesis
41	23	100.0	98	2	I57601	tubulin alpha chai
42	23	100.0	98	2	T36896	hypothetical prote
43	23	100.0	99	1	S30349	perchloric acid-so
44	23	100.0	99	2	T10839	embryonic abundant
45	23	100.0	100	2	C72655	hypothetical prote

ALIGNMENTS

```
RESULT  1
E61491
seed protein ws-5 - winged bean (fragment)
C;Species: Psophocarpus tetragonolobus (winged bean)
C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 07-Oct-1994
C;Accession: E61491
R;Hirano, H.
J. Protein Chem. 8, 115-130, 1989
A;Title: Microsequence analysis of winged bean seed proteins electroblotted from two-
A;Reference number: A61491; MUID:89351606
A;Accession: E61491
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-7 <HIR>
C;Keywords: glycoprotein; seed

Query Match             100.0%;  Score 23;  DB 2;  Length 7;
Best Local Similarity   100.0%;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VPGG 4
              ||||
Db          3 VPGG 6


RESULT  2
PC2215
fibrinogenolytic proteinase A2 (EC 3.4.21.-) - western diamondback rattlesnake (fragm
N;Alternate names: alpha-fibrinogenase A2
C;Species: Crotalus atrox (western diamondback rattlesnake)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Mar-1999
C;Accession: PC2215
R;Hung, C.C.; Chiou, S.H.
Biochem. Biophys. Res. Commun. 201, 1414-1423, 1994
A;Title: Isolation of multiple isoforms of alpha-fibrinogenase from the western diamo
viper.
A;Reference number: PC2214; MUID:94296418
A;Accession: PC2215
A;Molecule type: protein
A;Residues: 1-15 <HUN>
C;Keywords: hydrolase; serine proteinase

Query Match             100.0%;  Score 23;  DB 2;  Length 15;
Best Local Similarity   100.0%;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VPGG 4
              ||||
Db          1 VPGG 4
```

RESULT   3
A28391
hypothetical protein 1 - Streptomyces albidoflavus (fragment)
C;Species: Streptomyces albidoflavus
C;Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 11-Jan-2000
C;Accession: A28391
R;Long, C.M.; Virolle, M.J.; Chang, S.Y.; Chang, S.; Bibb, M.J.
J. Bacteriol. 169, 5745-5754, 1987
A;Title: Alpha-amylase gene of Streptomyces limosus: nucleotide sequence, expression mot[...]
A;Reference number: A91856; MUID:88058792
A;Accession: A28391
A;Molecule type: DNA
A;Residues: 1-27 <LON>
A;Note: the source is designated as Streptomyces limosus
C;Superfamily: lac repressor

Query Match   100.0%;   Score 23;   DB 2;   Length 27;
Best Local Similarity   100.0%;   Pred. No. 1.3e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
        |||
Db   24 VPGG 27

RESULT   4
S14062
hypothetical protein 1 - Streptomyces griseus (fragment)
C;Species: Streptomyces griseus
C;Date: 19-Mar-1997 #sequence_revision 26-Feb-1998 #text_change 11-Jan-2000
C;Accession: S14062
R;Vigal, T.; Gil, J.A.; Daza, A.; Garcia-Gonzalez, M.D.; Martin, J.F.
Mol. Gen. Genet. 225, 278-288, 1991
A;Title: Cloning, characterization and expression of an alpha-amylase gene from Strepton[...]
A;Reference number: S14062; MUID:91172128
A;Accession: S14062
A;Molecule type: DNA
A;Residues: 1-30 <VIG>
A;Experimental source: strain IMRU3570
C;Superfamily: lac repressor

Query Match   100.0%;   Score 23;   DB 2;   Length 30;
Best Local Similarity   100.0%;   Pred. No. 1.4e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
        |||
Db   27 VPGG 30

RESULT   5
I52232
tau protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C;Accession: I52232
R;Mori, H.; Hamada, Y.; Kawaguchi, M.; Honda, T.; Kondo, J.; Ihara, Y.
Biochem. Biophys. Res. Commun. 159, 1221-1226, 1989
A;Title: A distinct form of tau is selectively incorporated into Alzheimer's paired hel[...]
A;Reference number: I52232; MUID:89193714
A;Accession: I52232
A;Molecule type: mRNA
A;Residues: 1-31 <RES>
A;Cross-references: GB:M25299; NID:g602470; PIDN:AAA57264.1; PID:g602471
C;Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology

Query Match   100.0%;   Score 23;   DB 2;   Length 31;
Best Local Similarity   100.0%;   Pred. No. 1.4e+02;

Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
        |||
Db   26 VPGG 29

RESULT   6
T36863
probable two-component sensor kinase - Streptomyces coelicolor (fragment)
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36863
R;Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A[...]
submitted to the EMBL Data Library, June 1999
A;Reference number: Z2l616
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-41 <SEE>
A;Cross-references: EMBL:AL079332; PIDN:CAB45294.1; GSPDB:GN00070; SCOEDB:SC15.10c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC15.10c

Query Match   100.0%;   Score 23;   DB 2;   Length 41;
Best Local Similarity   100.0%;   Pred. No. 1.8e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
        |||
Db   38 VPGG 41

RESULT   7
S28397
T-complex protein 5 - bovine (fragments)
C;Species: Bos primigenius taurus (cattle)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Mar-1999
C;Accession: S28397
R;Frydman, J.; Nimmesgern, E.; Erdjument-Bromage, H.; Wall, J.S.; Tempst, P.; Hartl, [...]
EMBO J. 11, 4767-4778, 1992
A;Title: Function in protein folding of TRiC, a cytosolic ring complex containing TCP[...]
A;Reference number: S28395; MUID:93099850
A;Accession: S28397
A;Molecule type: protein
A;Residues: 1-20;21-45 <FRY>
C;Superfamily: molecular chaperone t-complex-type
C;Keywords: molecular chaperone

Query Match   100.0%;   Score 23;   DB 2;   Length 45;
Best Local Similarity   100.0%;   Pred. No. 2e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
        |||
Db   29 VPGG 32

RESULT   8
A61384
tracheal mucin glycoprotein - pig (fragments)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 09-Sep-1994
C;Accession: A61384
R;Sangadala, S.; Kim, D.; Brewer, J.M.; Mendicino, J.
Mol. Cell. Biochem. 102, 71-93, 1991
A;Title: Subunit structure of deglycosylated human and swine trachea and Cowper's gla[...]
A;Reference number: A61384; MUID:91270244
A;Accession: A61384
A;Status: preliminary

A;Molecule type: protein
A;Residues: 1-47 <SAN>
C;Keywords: glycoprotein

Query Match   100.0%;   Score 23;   DB 2;   Length 47;
Best Local Similarity 100.0%;   Pred. No. 2.1e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy   1 VPGG 4
       |||
Db  23 VPGG 26
```

RESULT 9
T36928
Hypothetical protein SCI7.32 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36928
R;Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A;Reference number: 221574
A;Accession: T36928
A;Status: Preliminary; translated from GB/EMBL/DDBJ
N;Molecule type: DNA
A;Residues: 1-56 <SEE>
A;Cross-references: EMBL:AL096743; PIDN:CAB46415.1; GSPDB:GN00070; SCOEDB:SCI7.32
C;Genetics:
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCI7.32

Query Match   100.0%;   Score 23;   DB 2;   Length 56;
Best Local Similarity 100.0%;   Pred. No. 2.4e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy   1 VPGG 4
       |||
Db   5 VPGG 8
```

RESULT 10
S03097
aerC protein - Aeromonas sobria
C;Species: Aeromonas sobria
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 08-Oct-1999
A;Accession: S03097
R;Husslein, V.; Huhle, B.; Jarchau, T.; Lurz, R.; Goebel, W.; Chakraborty, T.
Mol. Microbiol. 2, 507-517, 1988
A;Title: Nucleotide sequence and transcriptional analysis of the aerCaerA region of Aero
A;Reference number: S03097; MUID:89013892
A;Accession: S03097
N;Molecule type: DNA
A;Residues: 1-61 <HUS>
A;Cross-references: EMBL:X00559; NID:g39016; PIDN:CAA68641.1; PID:g39017
C;Genetics:
A;Gene: aerC

Query Match   100.0%;   Score 23;   DB 2;   Length 61;
Best Local Similarity 100.0%;   Pred. No. 2.6e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy   1 VPGG 4
       |||
Db  35 VPGG 38
```

RESULT 11
S37135
class II histocompatibility antigen DR beta-6 chain - rhesus macaque (fragment)
C;Species: Macaca mulatta (rhesus macaque)

C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 06-Jan-1995
C;Accession: S37135
R;Slierendregt, B.L.; Otting, N.; van Besouw, N.; Jonker, M.; Bontrop, R.E.
submitted to the EMBL Data Library, September 1993
A;Description: Expansion of rhesus macaque DRB regions by duplication.
A;Reference number: S37135
A;Accession: S37135
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-62 <SLI>
A;Cross-references: EMBL:Z26183

Query Match   100.0%;   Score 23;   DB 2;   Length 62;
Best Local Similarity 100.0%;   Pred. No. 2.7e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy   1 VPGG 4
       |||
Db  41 VPGG 44
```

RESULT 12
T04493
hypothetical protein F8P16.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
A;Accession: T04493
R;Bevan, M.; Brandt, P.; Dose, S.; Jarke, D.; Scharfe, M.; Schon, O.; Hoheisel, J.; M
submitted to the Protein Sequence Database, April 1998
A;Reference number: Z15375
A;Accession: T04493
A;Molecule type: DNA
A;Residues: 1-63 <BEV>
A;Cross-references: EMBL:AL021633
A;Experimental source: cultivar Columbia; BAC clone F8P16
C;Genetics:
A;Map position: 4
A;Note: F8P16.80

Query Match   100.0%;   Score 23;   DB 2;   Length 63;
Best Local Similarity 100.0%;   Pred. No. 2.7e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy   1 VPGG 4
       |||
Db   7 VPGG 10
```

RESULT 13
F49410
t-complex polypeptide 1 homolog (peak 4 fraction) - rabbit (fragments)
N;Alternate names: chaperonin homolog (peak 4)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 21-Sep-1994 #sequence_revision 18-Nov-1994 #text_change 04-Sep-1998
C;Accession: F49410
R;Rommelaere, H.; Van Troys, M.; Gao, Y.; Melki, R.; Cowan, N.J.; Vandekerckhove, J.;
Proc. Natl. Acad. Sci. U.S.A. 90, 11975-11979, 1993
A;Title: Eukaryotic cytosolic chaperonin contains t-complex polypeptide 1 and seven r
A;Reference number: A49410; MUID:94089752
A;Accession: F49410
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-69 <ROM>
A;Experimental source: reticulocyte
A;Note: sequence modified after extraction from NCBI backbone
A;Note: sequence extracted from NCBI backbone (NCBIP:141034, NCBIP:141036)
A;Superfamily: molecular chaperone t-complex-type

Query Match   100.0%;   Score 23;   DB 2;   Length 69;
Best Local Similarity 100.0%;   Pred. No. 3e+02;

```
Matches      4;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy      1 VPGG 4
          ||||
Db     39 VPGG 42
```

RESULT 14
T36407
hypothetical protein SCF34.08c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 19-May-2000
C;Accession: T36407
R;Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1999
A;Reference number: 221606
A;Accession: T36407
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-70 <SRN>
A;Cross-references: EMBL:AL109974; PIDN:CAB53319.1; GSPDB:GN00070; SCOEDB:SCF34.08c
A;Experimental source: strain A3(2)
A;Genetics:
C;Gene: SCOEDB:SCF34.08c
C;Superfamily: Escherichia coli hypothetical 8K protein (fes 3' region)

```
Query Match             100.0%;  Score 23;  DB 2;  Length 70;
Best Local Similarity   100.0%;  Pred. No. 3e+02;
Matches      4;  Conservative     4;  Mismatches     0;  Indels     0;  Gaps     0;

Qy      1 VPGG 4
          ||||
Db     32 VPGG 35
```

RESULT 15
B69958
hypothetical protein yqgY - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C;Accession: B69958
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cha
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fujita, M.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
        M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Leger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, J.; Sato, T.; Scanlon,
A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seroa
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033
A;Accession: B69958
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-81 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634723; GB:AL009126.1; PIDN:CAB14409.1; PID:g2634912
A;Experimental source: strain 168
C;Genetics:
A;Gene: yqgY

```
Query Match             100.0%;  Score 23;  DB 2;  Length 81;
Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
Matches      4;  Conservative     4;  Mismatches     0;  Indels     0;  Gaps     0;

Qy      1 VPGG 4
```

```
Db     75 VPGG 78
          ||||
```

Search completed: April 24, 2001, 16:42:04
Job time: 469 sec

Tue Apr 24 16:55:44 2001

```
                    GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        April 24, 2001, 16:42:50 ; Search time 44.88 Seconds
                                (without alignments)
                               3.053 Million cell updates/sec

Title:          US-09-340-736-7
Perfect score:  23
Sequence:       1 VPGG 4

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters:    93435

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SwissProt_39:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                            SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	23	100.0	23	1	NEU_LITCE	P81872 litoria cae
2	23	100.0	27	1	GRP_CANFA	P08969 canis famil
3	23	100.0	29	1	TLP_ACTDE	P81370 actinidia d
4	23	100.0	61	1	AERC_AERSO	P09165 aeromonas s
5	23	100.0	81	1	PN3C_PENVA	P81060 penaeus van
6	23	100.0	81	1	YOGY_BACSU	P54502 bacillus su
7	23	100.0	83	1	SEEP_RAPSA	P11573 raphanus sa
8	23	100.0	89	1	PTSN_PSEPU	P33670 pseudomonas
9	23	100.0	90	1	LSM5_HUMAN	Q9Y4Y9 homo sapien
10	23	100.0	91	1	EMB5_MAIZE	P45517 zea mays (m
11	23	100.0	91	1	EML_PICGL	Q40864 picea glauc
12	23	100.0	92	1	EM6_ARATH	Q02973 arabidopsis
13	23	100.0	92	1	EMB1_DAUCA	P17639 daucus caro
14	23	100.0	92	1	LE10_HELAN	P46514 helianthus
15	23	100.0	92	1	RL44_HALMA	P32411 haloarcula
16	23	100.0	93	1	EM1_WHEAT	P04568 triticum ae
17	23	100.0	93	1	EM3_WHEAT	Q08000 triticum ae
18	23	100.0	93	1	EM4_WHEAT	P47755 triticum ae
19	23	100.0	93	1	L19A_HORVU	C05190 hordeum vul
20	23	100.0	93	1	L19B_HORVU	P46532 hordeum vul
21	23	100.0	93	1	LSM5_YEAST	P40089 saccharomyc
22	23	100.0	94	1	RL44_ARCFU	O28936 archaeoglob
23	23	100.0	94	1	EM2_WHEAT	P22701 triticum ae
24	23	100.0	95	1	EMP1_ORYSA	P46520 oryza sativ
25	23	100.0	102	1	EM19_GOSHI	P09443 gossypium h
26	23	100.0	102	1	MGN_ORYSA	P49030 oryza sativ
27	23	100.0	111	1	SC4_SCHCO	P16934 schizophyll
28	23	100.0	112	1	YMFM_ECOLI	P75977 escherichia
29	23	100.0	116	1	YIS7_AQUAE	O67709 aquifex aeo
30	23	100.0	124	1	EP4_CANFA	Q28894 canis famil
31	23	100.0	125	1	Y4SN_RHISN	P50358 rhizobium s
32	23	100.0	126	1	TKNK_BOVIN	P08858 bos taurus
33	23	100.0	126	1	WN72_EPTST	P28123 eptatretus
34	23	100.0	130	1	FLHE_ECOLI	P76297 escherichia
35	23	100.0	130	1	FLHE_SALTY	P40728 salmonella
36	23	100.0	131	1	Y4IP_RHISN	P55459 rhizobium s
37	23	100.0	133	1	L193_HORVU	Q02400 hordeum vul
38	23	100.0	136	1	UK14_RAT	P52759 rattus norv
39	23	100.0	137	1	PA2C_MOUSE	P97391 mus musculu
40	23	100.0	137	1	PA2C_RAT	P51433 rattus norv
41	23	100.0	137	1	UK14_CAPHI	P80601 capra hircu
42	23	100.0	141	1	Y549_METJA	Q57969 methanococc
43	23	100.0	145	1	MT40_MYCTU	Q04001 mycobacteri
44	23	100.0	147	1	YV59_CABEL	P50439 caenorhabdi
45	23	100.0	152	1	EM1_ARATH	Q07187 arabidopsis

```
                            ALIGNMENTS

RESULT 1
NEU_LITCE
ID   NEU_LITCE       STANDARD;     PRT;      23 AA.
AC   P81872
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   NEUROMEDIN U-23 (NMU-23).
OS   Litoria caerulea.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;
OC   Litoria.
OX   NCBI_TaxID=30344;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Skin;
RX   MEDLINE=20138182; PubMed=10671478;
RA   Salmon A.L., Johnsen A.H., Bienert M., McMurray G., Nandha K.A.,
RA   Bloom S.R., Shaw C.;
RT   "Isolation, structural characterization and bioactivity of a novel
RT   neuromedin U analog from the defensive skin secretion of the
RT   Australasian tree frog Litoria caerulea.";
RL   J. Biol. Chem. 275:4549-4554(2000).
CC   -!- FUNCTION: STIMULATE UTERRINE SMOOTH MUSCLE CONTRACTION AND CAUSE
CC       SELECTIVE VASOCONSTRICTION.
CC   -!- SIMILARITY: BELONGS TO THE NMU FAMILY.
DR   InterPro; IPR001942; -.
KW   PROSITE; PS00967; NMU; 1.
KW   Amidation; Hormone.
FT   MOD_RES      23      23       AMIDATION.
SQ   SEQUENCE   23 AA;    2581 MW;  A9495841SCB58DC3 CRC64;

Query Match             100.0%;  Score 23;  DB 1;  Length 23;
Best Local Similarity   100.0%;  Pred. No. 63;
Matches    4;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy          1 VPGG 4
              ||||
Db          7 VPGG 10

RESULT 2
GRP_CANFA
ID   GRP_CANFA       STANDARD;     PRT;      27 AA.
AC   P08969
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GASTRIN-RELEASING PEPTIDE (GRP) [CONTAINS: NEUROMEDIN C (GRP-10)].
DE   GRP.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
```

RP   SEQUENCE.
RX   MEDLINE=83213518; PubMed=6853532;
RA   Reeve J.R. Jr., Walsh J.H., Chew P., Clark B., Hawke D.,
RA   Shively J.E.;
RT   "Amino acid sequences of three bombesin-like peptides from canine
RT   intestine extracts.";
RL   J. Biol. Chem. 258:5582-5588(1983).
CC   -!- FUNCTION: GRP STIMULATES GASTRIN RELEASE AS WELL AS OTHER
CC   GASTROINTESTINAL HORMONES.
CC   -!- SIMILARITY: BELONGS TO THE BOMBESIN/NEUROMEDIN B/RANATENSIN
CC   FAMILY.
DR   InterPro; IPR000874; -.
DR   Pfam; PF02044; Bombesin; 1.
DR   PROSITE; PS00257; BOMBESIN; 1.
KW   Bombesin family; Amidation.
FT   PEPTIDE   18   27
FT   MOD_RES   27   27   AMIDATION.
SQ   SEQUENCE   27 AA; 2889 MW; 9D9317261B7C1D65 CRC64;

Query Match   100.0%; Score 23; DB 1; Length 27;
Best Local Similarity   100.0%; Pred. No. 72;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
              ||||
Db   3 VPGG 6

RESULT 3
ID   TLP_ACTDE   STANDARD;   PRT;   29 AA.
AC   P81370;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   THAUMATIN-LIKE PROTEIN (FRAGMENT).
OS   Actinidia deliciosa (Kiwi).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae; Ericales;
OC   Actinidiaceae; Actinidia.
OX   NCBI_TaxID=3627;
RN   [1]
RP   SEQUENCE.
RC   STRAIN=CV. HAYWARD; TISSUE=Stem plug;
RA   Wurms K.V., Greenwood D.R., Sharrock K.R., Long P.G.;
RT   "Thaumatin-like protein in kiwifruit.";
RL   J. Sci. Food Agric. 79:1448-1452(1999).
CC   -!- TISSUE SPECIFICITY: WOODY STEM PLUG.
CC   -!- SIMILARITY: BELONGS TO THE THAUMATIN FAMILY.
     HSSP; P25871; 1AUN.
     InterPro; IPR001938; -.
DR   Pfam; PF00314; thaumatin; 1.
DR   PROSITE; PS00316; THAUMATIN; PARTIAL.
FT   NON_TER   29   29
SQ   SEQUENCE   29 AA; 3074 MW; E99F916F60AC3773 CRC64;

Query Match   100.0%; Score 23; DB 1; Length 29;
Best Local Similarity   100.0%; Pred. No. 77;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
              ||||
Db   18 VPGG 21

RESULT 4
AERC_AERSO
ID   AERC_AERSO   STANDARD;   PRT;   61 AA.
AC   P09165;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-MAR-1989 (Rel. 10, Last annotation update)
DE   AEROLYSIN REGULATORY PROTEIN.
GN   AERC.
OS   Aeromonas sobria.
OC   Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC   Aeromonas.
OX   NCBI_TaxID=646;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3;
RX   MEDLINE=89013892; PubMed=2459581;
RA   Husslein V., Huhle B., Jarchau T., Lurz R., Goebel W., Chakraborty T.;
RT   "Nucleotide sequence and transcriptional analysis of the aerCaerA
RT   region of Aeromonas sobria encoding aerolysin and its regulatory
RT   region.";
RL   Mol. Microbiol. 2:507-517(1988).
CC   -!- FUNCTION: REGULATION OF THE EXPRESSION OF AEROLYSIN.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Y00559; CAA68641.1; -.
DR   PIR; S03097; S03097.
SQ   SEQUENCE   61 AA; 6452 MW; DCFC7C7DBA7ED752 CRC64;

Query Match   100.0%; Score 23; DB 1; Length 61;
Best Local Similarity   100.0%; Pred. No. 1.5e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
              ||||
Db   35 VPGG 38

RESULT 5
PN3C_PENVA
ID   PN3C_PENVA   STANDARD;   PRT;   81 AA.
AC   P81060;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PENAEIDIN-3C PRECURSOR (P3-C).
OS   Penaeus vannamei (Penaeid shrimp) (European white shrimp).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
OC   Penaeidae; Penaeus.
OX   NCBI_TaxID=6689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Hemocyte;
RX   MEDLINE=98192209; PubMed=9353298;
RA   Destoumieux D., Bulet P., Loew D., van Dorsselaer A., Rodriguez J.,
RA   Bachere E.;
RT   "Penaeidins, a new family of antimicrobial peptides isolated from the
RT   shrimp Penaeus vannamei (Decapoda).";
RL   J. Biol. Chem. 272:28398-28406(1997).
CC   -!- FUNCTION: ANTIBACTERIAL AND ANTIMICROBIAL ACTIVITY.
CC   -!- TISSUE SPECIFICITY: HEMOCYTES.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --

```
DR   EMBL; Y14928; CAA75145.1; -.
KW   Antibiotic; Amidation; Signal.
FT   SIGNAL        1      ?       POTENTIAL.
FT   PROPEP        ?     19       BY SIMILARITY.
FT   CHAIN        20     80       PENAEIDIN-3C.
FT   MOD_RES      20     20       PYRROLIDONE CARBOXYLIC ACID (BY
FT                                SIMILARITY).
FT   MOD_RES      80     80       AMIDATION (G-81 PROVIDE AMIDE GROUP) (BY
FT                                SIMILARITY).
FT   DOMAIN       29     46       PRO-RICH.
SQ   SEQUENCE     81 AA;  8637 MW;  295083616 3B92C36 CRC64;

Query Match             100.0%;  Score 23;  DB 1;  Length 81;
Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 VPGG 4
            ||||
Db       39 VPGG 42


RESULT 6
ID   YQGI_BACSU     STANDARD;     PRT;    81 AA.
AC   P54502;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   HYPOTHETICAL 9.4 KDA PROTEIN IN SODA-COMGA INTERGENIC REGION.
GN   YQGI.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RA   Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA   Sato T., Takeuchi M.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; D84432; BAA12528.1; -.
DR   EMBL; Z99116; CAB14409.1; -.
DR   SubtiList; BG11692; YqgI.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     27     47       POTENTIAL.
FT   TRANSMEM     54     74       POTENTIAL.
SQ   SEQUENCE     81 AA;  9422 MW;  B22D7BF84B0DEA29 CRC64;

Query Match             100.0%;  Score 23;  DB 1;  Length 81;
Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 VPGG 4
            ||||
Db       75 VPGG 78


RESULT 7
ID   SEEP_RAPSA     STANDARD;     PRT;    83 AA.
AC   P15573;
DT   01-OCT-1989 (Rel. 12, Created)
```

```
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   LATE SEED MATURATION PROTEIN PSB6.
OS   Raphanus sativus (Radish).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Raphanus.
OX   NCBI_TaxID=3726;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. AVIGNON; TISSUE=Seed;
RA   Raynal X., Depigny D., Cooke R., Delseny M.;
RT   "Characterization of a radish nuclear gene expressed during late seed
RT   maturation.";
RL   Plant Physiol. 91:829-836(1989).
CC   -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN EQUIPPING THE SEED FOR
CC       SURVIVAL, MAINTAINING A MINIMAL LEVEL OF HYDRATION IN THE DRY
CC       ORGANISM AND PREVENTING THE DENATURATION OF CYTOPLASMIC
CC       COMPONENTS, OR MAY PLAY A ROLE DURING IMBIBITION BY CONTROLLING
CC       WATER UPTAKE.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE SEED MATURATION.
CC   -!- INDUCTION: BY ABSCISIC ACID (ABA) (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE SMALL HYDROPHILIC PLANT SEED PROTEIN
CC       FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X15440; CAA33479.1; -.
DR   EMBL; M31978; AAA33868.1; -.
DR   PIR; S04884; S04884.
DR   InterPro; IPR000389; -.
DR   Pfam; PF00477; seed_protein; 1.
DR   PROSITE; PS00431; SMALL_HYDR_PLANT_SEED; 1.
KW   Seed; Seed embryo.
SQ   SEQUENCE     83 AA;  8981 MW;  A930FB3F8C855172 CRC64;

Query Match             100.0%;  Score 23;  DB 1;  Length 83;
Best Local Similarity   100.0%;  Pred. No. 2e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 VPGG 4
            ||||
Db       21 VPGG 24


RESULT 8
ID   PTSN_PSEPU     STANDARD;     PRT;    89 AA.
AC   P33670;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   NITROGEN REGULATORY IIA PROTEIN (EC 2.7.1.69) (ENZYME IIA-NTR)
DE   (PHOSPHOTRANSFERASE ENZYME II, A COMPONENT?) (FRAGMENT).
GN   PTSN.
OS   Pseudomonas putida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN2100;
RX   MEDLINE=90152355; PubMed=2695395;
RA   Inouye S., Yamada M., Nakazawa A., Nakazawa T.;
RT   "Cloning and sequence analysis of the ntrA (rpoN) gene of Pseudomonas
```

RT   putida.";
RL   Gene 85:145-152(1989).
CC   -!- FUNCTION: SEEMS TO HAVE A ROLE IN LINKING CARBON AND NITROGEN
CC   ASSIMILATION (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC   PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE PTS IIA FAMILY.
CC   --

DR   EMBL; M24916; -; NOT_ANNOTATED_CDS.
DR   HSSP; P31222; 1A6J.
DR   InterPro; IPR002178; -.
DR   Pfam; PF00359; PTS_EIIA_2; 1.
DR   PROSITE; PS00372; PTS_EIIA_2; 1.
KW   Phosphotransferase system; Transferase; Phosphorylation.
FT   MOD_RES   68   68   PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   89 AA;   9488 MW; 8CAA4C53914E1DD3 CRC64;

Query Match   100.0%; Score 23; DB 1; Length 89;
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches   4; Conservative   0; Mismatches   0;

Qy 1 VPGG 4
       ||||
Db 17 VPGG 20

RESULT 9
ID   LSM5_HUMAN   STANDARD;   PRT;   90 AA.
AC   Q9Y4Y9;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   U6 SNRNA-ASSOCIATED SM-LIKE PROTEIN LSM5.
GN   LSM5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=99298196; PubMed=10369684;
RA   Salgado-Garrido J., Bragado-Nilsson E., Kandels-Lewis S., Seraphin B.;
RT   "Sm and Sm-like proteins assemble in two related complexes of deep
RT   evolutionary origin.";
RL   EMBO J. 18:3451-3462(1999).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=99452783; PubMed=10523320;
RA   Achsel T., Brahms H., Kastner B., Bachi A., Wilm M., Luehrmann R.;
RT   "A doughnut-shaped heteromer of human Sm-like proteins binds to the
RT   3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in
RT   vitro.";
RL   EMBO J. 18:5789-5802(1999).
CC   -!- FUNCTION: PLAYS A ROLE IN U6 SNRNP ASSEMBLY AND FUNCTION. BINDS TO
CC   THE 3' END OF U6 SNRNA, THEREBY FACILITATING U4/U6 DUPLEX
CC   FORMATION IN VITRO.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SNRNP SM PROTEINS FAMILY.
CC   --

DR   EMBL; AJ238097; CAB45868.1; -.
DR   EMBL; AF182291; AAD56229.1; -.
DR   InterPro; IPR001163; -.
DR   Pfam; PF01423; Sm; 1.
KW   Nuclear protein; Ribonucleoprotein; mRNA splicing; mRNA processing;
KW   RNA-binding; Acetylation.
FT   INIT_MET   0   0
FT   MOD_RES   1   1   ACETYLATION.
SQ   SEQUENCE   90 AA;   9806 MW; 4B45811B47B054DB CRC64;

Query Match   100.0%; Score 23; DB 1; Length 90;
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches   4; Conservative   0; Mismatches   0;

Qy 1 VPGG 4
       ||||
Db 82 VPGG 85

RESULT 10
ID   EMB5_MAIZE   STANDARD;   PRT;   91 AA.
AC   P46517;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   LATE EMBRYOGENESIS ABUNDANT PROTEIN EMB564.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
OC   Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. CO255; TISSUE=Embryo;
RX   MEDLINE=91316231; PubMed=1830496;
RA   Williams B., Tsang A.;
RT   "A maize gene expressed during embryogenesis is abscisic
RT   acid-inducible and highly conserved.";
RL   Plant Mol. Biol. 16:919-923(1991).
CC   -!- FUNCTION: LEA PROTEINS ARE LATE EMBRYONIC PROTEINS ABUNDANT IN
CC   HIGHER PLANT SEED EMBRYOS. THEY MAY PLAY AN ESSENTIAL ROLE IN
CC   SEED SURVIVAL AND IN CONTROLLING WATER EXCHANGES DURING SEED
CC   DESICCATION AND IMBIBITION.
CC   -!- INDUCTION: BY ABSCISIC ACID AND OSMOTIC STRESS.
CC   -!- SIMILARITY: BELONGS TO THE SMALL HYDROPHILIC PLANT SEED PROTEIN
CC   FAMILY.
CC   --

DR   EMBL; X55388; CAA39063.1; -.
DR   PIR; S16249; S16249.
DR   MaizeDB; 120720; -.
DR   InterPro; IPR000389; -.
DR   Pfam; PF00477; seed_protein; 1.
DR   PROSITE; PS00431; SMALL_HYDR_PLANT_SEED; 1.
KW   Seed embryo; Seed; Multigene family.
SQ   SEQUENCE   91 AA;   9684 MW; C7851750OA7F675C CRC64;

```
Query Match           100.0%;  Score 23;  DB 1;  Length 91;
Best Local Similarity 100.0%;  Pred. No. 2.2e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 VPGG 4
         |||
Db    22 VPGG 25

RESULT 11
ID   EML_PICGL      STANDARD;      PRT;      91 AA.
AC   Q40864;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   EM-LIKE PROTEIN.
OS   Picea glauca (White spruce).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Coniferopsida; Coniferales; Pinaceae; Picea.
OX   NCBI_TaxID=3330;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dong J.Z., Dunstan D.I.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE SMALL HYDROPHILIC PLANT SEED PROTEIN
CC       FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L42464; AAA85366.1; -.
DR   InterPro; IPR000389; -.
DR   Pfam; PF00477; seed_protein; 1.
DR   PROSITE; PS00431; SMALL_HYDR_PLANT_SEED; 1.
KW   Seed; Seed embryo.
SQ   SEQUENCE   91 AA;   9771 MW;   8FD9D4F41E0A195E CRC64;


Query Match           100.0%;  Score 23;  DB 1;  Length 91;
Best Local Similarity 100.0%;  Pred. No. 2.2e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 VPGG 4
         |||
Db    21 VPGG 24

RESULT 12
ID   EM6_ARATH      STANDARD;      PRT;      92 AA.
AC   Q02973;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   EM-LIKE PROTEIN GEA6.
GN   EM6 OR ATEM6 OR D1H OR AT2G40170.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=93261424; PubMed=8492809;
RA   Gaubier P., Raynal M., Hull G., Huestis G.M., Grellet F.,
RA   Arenas C., Pages M., Delseny M.;
RT   "Two different Em-like genes are expressed in Arabidopsis thaliana
RT   seeds during maturation.";
RL   Mol. Gen. Genet. 238:409-418(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. LANDSBERG ERECTA;
RX   MEDLINE=93261423; PubMed=8492808;
RA   Finkelstein R.R.;
RT   "Abscisic acid-insensitive mutations provide evidence for stage-
RT   specific signal pathways regulating expression of an Arabidopsis late
RT   embryogenesis-abundant (lea) gene.";
RL   Mol. Gen. Genet. 238:401-408(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99225673; PubMed=10207155;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
CC   -!- FUNCTION: IT IS THOUGHT TO PROVIDE PROTECTION FOR THE CYTOPLASM
CC       DURING THE DESICCATION STAGE OF EMBRYO DEVELOPMENT.
CC   -!- TISSUE SPECIFICITY: PRESENT ONLY IN NEARLY DRY AND DRY SEEDS.
CC   -!- INDUCTION: BY ABSCISIC ACID (ABA).
CC   -!- SIMILARITY: BELONGS TO THE SMALL HYDROPHILIC PLANT SEED PROTEIN
CC       FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z11923; CAA77981.1; -.
DR   EMBL; Z11924; CAA77982.1; -.
DR   EMBL; X66023; CAA46821.1; -.
DR   EMBL; AF085279; AAD25932.1; -.
DR   EMBL; AC018721; AAF18731.1; -.
DR   InterPro; IPR000389; -.
DR   Pfam; PF00477; seed_protein; 1.
DR   PROSITE; PS00431; SMALL_HYDR_PLANT_SEED; 1.
KW   Seed; Seed embryo; Multigene family.
SQ   SEQUENCE   92 AA;   9934 MW;   D35A7879AF773706 CRC64;


Query Match           100.0%;  Score 23;  DB 1;  Length 92;
Best Local Similarity 100.0%;  Pred. No. 2.2e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1 VPGG 4
         |||
Db    21 VPGG 24

RESULT 13
ID   EMB1_DAUCA      STANDARD;      PRT;      92 AA.
```

```
AC   P17639;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   EMB-1 PROTEIN.
GN   EMB-1.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
OC   euasterids II; Apiales; Apiaceae; Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DANVERS; TISSUE=Embryo;
RX   MEDLINE=90251475; PubMed=2339072;
RA   Ulrich T.H., Wurtele E.S., Nikolau B.J.;
RT   "Sequence of EMB-1, an mRNA accumulating specifically in embryos of
RT   carrot.";
RL   Nucleic Acids Res. 18:2826-2826(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. DANVERS; TISSUE=Embryo;
RX   MEDLINE=94151426; PubMed=8108498;
RA   Wurtele E.S., Wang H., Durgerian S., Nikolau B.J., Ulrich T.H.;
RT   "Characterization of a gene that is expressed early in somatic
RT   embryogenesis of Daucus carota.";
RL   Plant Physiol. 102:303-312(1993).
CC   -!- DEVELOPMENTAL STAGE: EMBRYOS.
CC   -!- SIMILARITY: BELONGS TO THE SMALL HYDROPHILIC PLANT SEED PROTEIN
CC       FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X60131; CAA42717.1; -.
DR   EMBL; X17608; CAA35610.1; -.
DR   PIR; S10010; S10010.
DR   PIR; JQ2273; JQ2273.
DR   InterPro; IPR000389; -.
DR   Pfam; PF00477; seed_protein; 1.
DR   PROSITE; PS00431; SMALL_HYDR_PLANT_SEED; 1.
KW   Seed; Seed embryo.
SQ   SEQUENCE 92 AA; 9917 MW; 13E09FA58F5F4FFA CRC64;

Query Match    100.0%;   Score 23;   DB 1;   Length 92;
Best Local Similarity  100.0%;   Pred. No. 2.2e+02;
Matches    4;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy       1 VPGG 4
           ||||
Db      21 VPGG 24

RESULT 14
ID   LE10_HELAN    STANDARD;    PRT;    92 AA.
AC   P46514;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   10 KDA LATE EMBRYOGENESIS ABUNDANT PROTEIN (DS10).
OS   Helianthus annuus (Common sunflower).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
OC   euasterids II; Asterales; Asteraceae; Asteroideae; Heliantheae;
OC   Helianthus.
OX   NCBI_TaxID=4232;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. SUNWEED; TISSUE=Dry seed;
RX   MEDLINE=92353387; PubMed=1386536;
RA   Almoguera C., Jordano J.;
RT   "Developmental and environmental concurrent expression of sunflower
RT   dry-seed-stored low-molecular-weight heat-shock protein and Lea
RT   mRNAs.";
RL   Plant Mol. Biol. 19:781-792(1992).
RN   [2]
RP   REVISION TO 51.
RC   STRAIN=CV. SUNWEED; TISSUE=Cotyledon;
RA   Jordano J.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99190608; PubMed=10092187;
RA   Prieto-Dapena P., Almoguera C., Rojas A., Jordano J.;
RT   "Seed-specific expression patterns and regulation by ABI3 of an
RT   unusual late embryogenesis abundant gene in sunflower.";
RL   Plant Mol. Biol. 39:615-627(1999).
CC   -!- FUNCTION: LEA PROTEINS ARE LATE EMBRYONIC PROTEINS ABUNDANT IN
CC       HIGHER PLANT SEED EMBRYOS. THEY MAY PLAY AN ESSENTIAL ROLE IN
CC       SEED SURVIVAL AND IN CONTROLLING WATER EXCHANGES DURING SEED
CC       DESICCATION AND IMBIBITION.
CC   -!- TISSUE SPECIFICITY: MAXIMALLY EXPRESSED IN DRY SEEDS. ALSO
CC       PRESENT IN MID-MATURATION EMBRYOS.
CC   -!- INDUCTION: BY ABSCISIC ACID. OSMOTIC STRESS AND HEAT SHOCK.
CC   -!- SIMILARITY: BELONGS TO THE SMALL HYDROPHILIC PLANT SEED PROTEIN
CC       FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X59699; CAA42220.1; -.
DR   EMBL; AJ224115; CAA11834.1; -.
DR   InterPro; IPR000389; -.
DR   Pfam; PF00477; seed_protein; 1.
DR   PROSITE; PS00431; SMALL_HYDR_PLANT_SEED; 1.
KW   Seed embryo; Seed.
SQ   SEQUENCE 92 AA; 10037 MW; BD612FA4CE72BAFF CRC64;

Query Match    100.0%;   Score 23;   DB 1;   Length 92;
Best Local Similarity  100.0%;   Pred. No. 2.2e+02;
Matches    4;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy       1 VPGG 4
           ||||
Db      31 VPGG 34

RESULT 15
ID   RL44_HALMA    STANDARD;    PRT;    92 AA.
AC   P32411;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   50S RIBOSOMAL PROTEIN L44E (HL44).
GN   RPL44E.
OS   Haloarcula marismortui (Halobacterium marismortui).
OC   Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloarcula.
OX   NCBI_TaxID=2238;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=93277953; PubMed=8504167;
RA   Bergmann U., Wittmann-Liebold B.;
```

RT   "HL35e and HLA: primary structure of two very basic and cysteine-rich
RT   ribosomal proteins from Haloarcula marismortui.";
RL   Biochim. Biophys. Acta 1173:195-200(1993).
CC   -!- SIMILARITY: BELONGS TO THE L44E FAMILY OF RIBOSOMAL PROTEINS.
DR   PIR; S33790; S33790.
DR   InterPro; IPR000552; -.
DR   Pfam; PF00935; Ribosomal_L44; 1.
DR   PROSITE; PS01172; RIBOSOMAL_L44E; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   92 AA; 10791 MW; 14197AA6AE7AC07A CRC64;

Query Match   100.0%; Score 23; DB 1; Length 92;
Best Local Similarity   100.0%; Pred. No. 2.2e+02;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
                |||
Db   55 VPGG 58

Search completed: April 24, 2001, 16:42:52
Job time: 458 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:40:37 ; Search time 125.5 seconds
(without alignments)
3.736 Million cell updates/sec

Title: US-09-340-736-7
Perfect score: 23
Sequence: 1 VPGG 4

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   SPTREMBL_15:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match	Length	DB	ID	Description
1	23	100.0	23	15	13	Q9PRW3	q9prw3 crotalus at
2	23	100.0	23	20	2	Q46499	q46499 desulfovibr
3	23	100.0	23	27	14	O89142	o89142 hepatitis c
4	23	100.0	23	28	10	Q9S886	q9s886 citrus sine
5	23	100.0	23	30	6	Q9N2H3	q9n2h3 bos taurus
6	23	100.0	23	31	4	Q9UMH0	q9umh0 homo sapien
7	23	100.0	23	41	2	Q9X9J0	q9x9j0 streptomyce
8	23	100.0	23	44	2	Q9ZFC8	q9zfc8 lactococcus
9	23	100.0	23	44	6	Q9TSD3	q9tsd3 oryctolagus
10	23	100.0	23	44	8	Q9T2F3	q9t2f3 brassica na
11	23	100.0	23	45	6	Q9TQR1	q9tqr1 bos taurus
12	23	100.0	23	51	10	Q9SXF3	q9sxf3 gossypium s
13	23	100.0	23	56	7	Q9X9Y8	q9x9y8 streptomyce
14	23	100.0	23	57	5	Q29978	q29978 homo sapien
15	23	100.0	23	61	6	Q53455	q53455 streptomyce
16	23	100.0	23	63	10	O49576	o49576 arabidopsis
17	23	100.0	23	67	2	Q9L8X0	q9l8x0 streptomyce
18	23	100.0	23	70	2	Q9RK17	q9rk17 streptomyce
19	23	100.0	23	70	2	Q52113	q52113 acinetobact
20	23	100.0	23	73	10	Q41938	arabidopsis
21	23	100.0	23	73	10	Q9M5K8	zea mays (m
22	23	100.0	23	74	2	Q9L5Q9	salmonella
23	23	100.0	23	76	5	Q9P2Z3	homo sapien
24	23	100.0	23	78	5	Q9VVN7	drosophila
25	23	100.0	23	81	2	P94260	borrelia bu
26	23	100.0	23	81	2	P94261	borrelia bu
27	23	100.0	23	81	2	P94262	borrelia bu
28	23	100.0	23	83	11	P97880	rattus norv
29	23	100.0	23	84	10	O65725	brassica na
30	23	100.0	23	84	11	O60471	cavia porce
31	23	100.0	23	89	5	O15348	homo sapien
32	23	100.0	23	90	6	O77614	ovis aries
33	23	100.0	23	91	4	Q9RY58	deinococcus
34	23	100.0	23	91	4	O15349	homo sapien
35	23	100.0	23	91	4	Q9Y4Y9	homo sapien
36	23	100.0	23	91	5	O18694	caenorhabdi
37	23	100.0	23	91	5	Q9VRT7	drosophila
38	23	100.0	23	91	10	O40860	picea glauc
39	23	100.0	23	92	10	Q39095	arabidopsis
40	23	100.0	23	93	1	Q9YEC1	aeropyrum p
41	23	100.0	23	93	5	Q18842	caenorhabdi
42	23	100.0	23	93	10	Q9ZR70	triticum ae
43	23	100.0	23	93	10	Q9M4P1	secale cere
44	23	100.0	23	93	10	Q9M4P0	secale cere
45	23	100.0	23	93	10	Q9M4N9	secale cere

ALIGNMENTS

RESULT 1
Q9PRW3
ID   Q9PRW3   PRELIMINARY;   PRT;   15 AA.
AC   Q9PRW3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ALPHA-FIBRINOGENASE ISOFORM A2 (FRAGMENT).
OS   Crotalus atrox (Western diamondback rattlesnake).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Crotalus.
OX   NCBI_TaxID=8730;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=94296418; PubMed=8024586;
RA   Hung C.C., Chiou S.H.;
RT   "Isolation of multiple isoforms of alpha-fibrinogenase from the
RT   Western diamondback rattlesnake, Crotalus atrox: N-terminal sequence
RT   homology with ancrod, an antithrombotic agent from Malayan viper.";
RL   Biochem. Biophys. Res. Commun. 201:1414-1423(1994).
SQ   SEQUENCE   15 AA; 1640 MW; 03EFE10227CA12DA CRC64;

Query Match   100.0%; Score 23; DB 13; Length 15;
Best Local Similarity   100.0%; Pred. No. 99;
Matches   4; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VPGG 4
          ||||
Db   1 VPGG 4

RESULT 2
Q46499
ID   Q46499   PRELIMINARY;   PRT;   20 AA.
AC   Q46499;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE   PERIPLASMIC (FRAGMENT).
OS   Desulfovibrio desulfuricans.

OC   Bacteria; Proteobacteria; delta subdivision; Desulfovibrio.
OX   NCBI_TaxID=876;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M6;
RA   Lim S.K., Lee S.J., Kim B.H.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U49192; AAA91808.1; -.
FT   NON_TER   1   1
FT   NON_TER   20   20
SQ   SEQUENCE   20 AA;   1996 MW;   E93207D26C22999B CRC64;

Query Match   100.0%;   Score 23;   DB 2;   Length 20;
Best Local Similarity   100.0%;   Pred. No. 1.3e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
     ||||
Db   3 VPGG 6

RESULT 3
AC   089142;
ID   089142;   PRELIMINARY;   PRT;   27 AA.
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   E2 GLYCOPROTEIN (FRAGMENT).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1B-2;
RX   MEDLINE=98378035; PubMed=9714233;
RA   Kato N., Ikeda M., Sugiyama K., Mizutani T., Tanaka T., Shimotohno K.;
RT   "Hepatitis C virus population dynamics in human lymphocytes and
RT   hepatocytes infected in vitro.";
RL   J. Gen. Virol. 79:1859-1869(1998).
DR   EMBL; AB007716; BAA32637.1; -.
DR   INTERPRO; IPR002531; -.
DR   PFAM; PF01560; HCV_NS1; 1.
FT   NON_TER   1   1
FT   NON_TER   27   27
SQ   SEQUENCE   27 AA;   2861 MW;   BB1F4D9A2AB04A25 CRC64;

Query Match   100.0%;   Score 23;   DB 14;   Length 27;
Best Local Similarity   100.0%;   Pred. No. 1.8e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
     ||||
Db   4 VPGG 7

RESULT 4
ID   Q9S886;
AC   Q9S886;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BASIC CHITINASE (EC 3.2.1.14) (FRAGMENT).
OS   Citrus sinensis (Sweet orange).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Sapindales; Rutaceae; Citrus.
OX   NCBI_TaxID=2711;
RN   [1]
ID   Q9S886;   PRELIMINARY;   PRT;   28 AA.

RP   SEQUENCE.
RX   MEDLINE=97085076; PubMed=8931350;
RA   Mayer R.T., McCollum T.G., Niedz R.P., Hearn C.J., McDonald R.E.,
RA   Berdis E., Doostdar H.;
RT   "Characterization of seven basic endochitinases isolated from cell
RT   cultures of Citrus sinensis (L.).";
RL   Planta 200:289-295(1996).
DR   HSSP; P25871; 1AUN.
DR   INTERPRO; IPR001938; -.
DR   PFAM; PF00314; thaumatin; 1.
DR   PRINTS; PR00347; THAUMATIN.
SQ   SEQUENCE   28 AA;   3041 MW;   FE27A9A1AD305AF5 CRC64;

Query Match   100.0%;   Score 23;   DB 10;   Length 28;
Best Local Similarity   100.0%;   Pred. No. 1.9e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
     ||||
Db   18 VPGG 21

RESULT 5
Q9N2H3
ID   Q9N2H3;   PRELIMINARY;   PRT;   30 AA.
AC   Q9N2H3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BCNT.
DE   BUCENTAUR (FRAGMENT).
GN   BCNT.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Iwashita S., Sezaki M., Takahashi I., Takeda H., Suginoto Y.;
RT   "Gene Organization of bovine bcnt that contains LINE-derived region in
RT   its protein.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Takahashi I., Nobukuni T., Ohmori H., Kobayashi M., Tanaka S.,
RA   Ohshima K., Okada M., Masui T., Hashimoto K., Iwashita S.;
RT   "Existence of a bovine LINE repetitive insert that appears in the cDNA
RT   of bovine protein BCNT in ruminant, but not in human, genomes.";
RL   Gene 211:387-394(1998).
DR   EMBL; AB033999; BAA93548.1; -.
FT   NON_TER   1   1
FT   NON_TER   30   30
SQ   SEQUENCE   30 AA;   2951 MW;   D32134B9AF2A332F CRC64;

Query Match   100.0%;   Score 23;   DB 6;   Length 30;
Best Local Similarity   100.0%;   Pred. No. 2.1e+02;
Matches   4;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 VPGG 4
     ||||
Db   11 VPGG 14

RESULT 6
Q9UMH0
ID   Q9UMH0;   PRELIMINARY;   PRT;   31 AA.
AC   Q9UMH0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   TAU PROTEIN (FRAGMENT).

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=AGED BRAIN;
RX   MEDLINE=89193714; PubMed=2495000;
RA   Mori H., Hamada Y., Kawaguchi M., Honda T., Kondo J., Ihara Y.;
RT   "A distinct form of tau is selectively incorporated into Alzheimer's
RT   paired helical filaments.";
RL   Biochem. Biophys. Res. Commun. 159:1221-1226(1989).
DR   EMBL; M25298; AAA57264.1; -.
DR   INTERPRO; IPR001084; -.
DR   PFAM; PF00418; tubulin-binding; 1.
DR   PROSITE; PS00229; TAU_MAP; 1.
GN
FT   NON_TER     1      1
FT   NON_TER    31     31
SQ   SEQUENCE   31 AA;  3265 MW;  5CG2909A643A1C9E CRC64;

Query Match             100.0%;  Score 23;  DB 4;  Length 31;
Best Local Similarity   100.0%;  Pred. No. 2.1e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 VPGG 4
          ||||
Db   26 VPGG 29

RESULT 7
Q9SX920
ID   Q9SX920           PRELIMINARY;     PRT;    41 AA.
AC   Q9SX920;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   PUTATIVE TWO COMPONENT SENSOR KINASE (FRAGMENT).
GN   SCI5.10C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K., Harris D.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
     EMBL; AL079333; CAA45294.1; -.
DR   INTERPRO; IPR000410; -.
DR   PRINTS; PR00344; BCTRLSENSOR.
KW   Kinase.
FT   NON_TER     1      1
SQ   SEQUENCE   41 AA;  4221 MW;  ABCC81778A7A98A0D CRC64;

Query Match             100.0%;  Score 23;  DB 2;  Length 41;
```

```
Best Local Similarity   100.0%;  Pred. No. 2.8e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 VPGG 4
          ||||
Db   38 VPGG 41

RESULT 8
Q92FC8
ID   Q92FC8            PRELIMINARY;     PRT;    44 AA.
AC   Q92FC8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   THIOREDOXIN REDUCTASE (FRAGMENT).
GN   TRXB.
OS   Lactococcus lactis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1358;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL1403;
RA   Aleksandrzak T., Kowalczyk M., Kok J., Bardowski J.;
RT   "Dual role of CcpA protein in regulation of sugar metabolism in
RT   Lactococcus lactis.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF106673; AAC96331.1; -.
DR   INTERPRO; IPR000103; -.
DR   INTERPRO; IPR000447; -.
DR   INTERPRO; IPR000759; -.
DR   PRINTS; PR00419; ADXRDTASE.
DR   PRINTS; PR00469; PNDRDTASEII.
DR   PRINTS; PR01001; FADG3PDH.
FT   NON_TER    44     44
SQ   SEQUENCE   44 AA;  4503 MW;  5B9AF8B3C1411A0C CRC64;

Query Match             100.0%;  Score 23;  DB 2;  Length 44;
Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 VPGG 4
          ||||
Db   37 VPGG 40

RESULT 9
Q9TSD3
ID   Q9TSD3            PRELIMINARY;     PRT;    44 AA.
AC   Q9TSD3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CHAPERONIN-T-COMPLEX POLYPEPTIDE 1 HOMOLOG (FRAGMENT).
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94089752; PubMed=7903455;
RA   Rommelaere H., Van Troys M., Gao Y., Melki R., Cowan N.J.,
RA   Vandekerckhove J., Ampe C.;
RT   "Eukaryotic cytosolic chaperonin contains t-complex polypeptide 1 and
RT   seven related subunits.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:11975-11979(1993).
FT   NON_TER     1      1
FT   NON_CONS   18     19
FT   NON_TER    44     44
SQ   SEQUENCE   44 AA;  4458 MW;  E6391D615DD0ODF2 CRC64;
```

Left column

Query Match　100.0%;　Score 23;　DB 5;　Length 44;
Best Local Similarity　100.0%;　Pred. No. 3.1e+02;
Matches　4;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　1　VPGG　4
　　　||||
Db　27　VPGG　30

RESULT 10
Q9T2T3
ID　Q9T2T3　PRELIMINARY;　PRT;　44 AA.
AC　Q9T2T3;
DT　01-MAY-2000 (TrEMBLrel. 13, Created)
DT　01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT　01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE　CHAPERONIN-60 LS2 FRAGMENT (FRAGMENTS).
OS　Brassica napus (Rape).
OC　Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC　Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC　Brassicales; Brassicaceae; Brassica.
OX　NCBI_TaxID=3708;
RN　[1]
RP　SEQUENCE.
RX　MEDLINE=94302168; PubMed=7913238;
RA　Cloney L.P., Bekkaoui D.R., Feist G.L., Lane W.S., Hemmingsen S.M.;
RT　"Brassica napus plastid and mitochondrial chaperonin-60 proteins
RT　contain multiple distinct polypeptides.";
RL　Plant Physiol. 105:233-241(1994).
DR　HSSP; P06139; 1DER.
DR　INTERPRO; IPR001844; -.
DR　INTERPRO; IPR002423; -.
DR　PFAM; PF00118; cpn60_TCP1; 1.
DR　PROSITE; PS00296; CHAPERONINS_CPN60; 1.
FT　NON_TER　1　1
FT　NON_TER　44　44
FT　NON_CONS　23　24
SQ　SEQUENCE　44 AA;　4541 MW;　8443D0C9462AF4F3 CRC64;

Query Match　100.0%;　Score 23;　DB 8;　Length 44;
Best Local Similarity　100.0%;　Pred. No. 3.1e+02;
Matches　4;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　1　VPGG　4
　　　||||
Db　32　VPGG　35

RESULT 11
Q9TQR1
ID　Q9TQR1　PRELIMINARY;　PRT;　45 AA.
AC　Q9TQR1;
DT　01-MAY-2000 (TrEMBLrel. 13, Created)
DT　01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT　01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE　CHAPERONIN (FRAGMENTS).
OS　Bos taurus (Bovine).
OC　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC　Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC　Bovidae; Bovinae; Bos.
OX　NCBI_TaxID=9913;
RN　[1]
RP　SEQUENCE.
RX　MEDLINE=94089752; PubMed=7903455;
RA　Ronnelaere H., Van Troys M., Gao Y., Melki R., Cowan N.J.,
RA　Vandekerckhove J., Ampe C.;
RT　"Eukaryotic cytosolic chaperonin contains t-complex polypeptide 1 and
RT　seven related subunits.";
RL　Proc. Natl. Acad. Sci. U.S.A. 90:11975-11979(1993).
RN　[2]

Right column

RP　SEQUENCE.
RX　MEDLINE=93099850; PubMed=1361170;
RA　Frydman J., Nimmesgern E., Erdjument-Bromage H., Wall J.S., Tempst P.,
RA　Hartl F.U.;
RT　"Function in protein folding of TRiC, a cytosolic ring complex
RT　containing TCP-1 and structurally related subunits.";
RL　EMBO J. 11:4767-4778(1992).
DR　INTERPRO; IPR002423; -.
DR　PFAM; PF00118; cpn60_TCP1; 1.
FT　NON_TER　20　21
FT　NON_CONS　20　21
FT　NON_TER　45　45
SQ　SEQUENCE　45 AA;　4962 MW;　1573A84E23BCDC9 CRC64;

Query Match　100.0%;　Score 23;　DB 5;　Length 45;
Best Local Similarity　100.0%;　Pred. No. 3.1e+02;
Matches　4;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　1　VPGG　4
　　　||||
Db　29　VPGG　32

RESULT 12
Q9SMF3
ID　Q9SMF3　PRELIMINARY;　PRT;　51 AA.
AC　Q9SMF3;
DT　01-MAY-2000 (TrEMBLrel. 13, Created)
DT　01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT　01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE　MICROSOMAL OMEGA6 DESATURASE ENZYME (FRAGMENT).
GN　FAD2.
OS　Gossypium stocksii.
OC　Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC　Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC　Malvales; Malvaceae; Gossypium.
OX　NCBI_TaxID=47602;
RN　[1]
RP　SEQUENCE FROM N.A.
RA　Brubaker C.L., Liu Q., Green A.G., Marshall D.R., Singh S.P.;
RT　"Microsomal omega6 desaturase intron topologies contribute to our
RT　understanding of reticulate evolution in Gossypium (Malvaceae) and the
RT　evolution of the Australian Gossypium species.";
RL　Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR　EMBL; AJ244917; CAB59280.1; -.
FT　NON_TER　1　1
FT　CHAIN　1　>51
FT　NON_TER　51　51
SQ　SEQUENCE　51 AA;　5728 MW;　B3103EFF7594BBA3 CRC64;

Query Match　100.0%;　Score 23;　DB 10;　Length 51;
Best Local Similarity　100.0%;　Pred. No. 3.6e+02;
Matches　4;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Qy　1　VPGG　4
　　　||||
Db　2　VPGG　5

RESULT 13
Q9X9V8
ID　Q9X9V8　PRELIMINARY;　PRT;　56 AA.
AC　Q9X9V8;
DT　01-NOV-1999 (TrEMBLrel. 12, Created)
DT　01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT　01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE　HYPOTHETICAL 6.4 KDA PROTEIN.
GN　SCI7.32.
OS　Streptomyces coelicolor.
OC　Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC　Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

```
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K., Harris D.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL096743; CAB46415.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   56 AA;   6392 MW;   CBE7F4D6BF0413B8 CRC64;

Query Match             100.0%;  Score 23;  DB 2;  Length 56;
Best Local Similarity   100.0%;  Pred. No. 3.9e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1  VPGG 4
              ||||
Db         5  VPGG 8


RESULT 14
Q29978
ID   Q29978          PRELIMINARY;      PRT;     57 AA.
AC   Q29978;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE   LYMPHOCYTE ANTIGEN (FRAGMENT).
DE   HLA-DQB1
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Meyer C.G.;
RL   Tissue Antigens 0:0-0(0).
DR   EMBL; M86226; AAA59692.1; -.
KW   MHC.
FT   NON_TER       1      1
SQ   SEQUENCE   57 AA;   5594 MW;   DD964A069939393C73 CRC64;

Query Match             100.0%;  Score 23;  DB 7;  Length 57;
Best Local Similarity   100.0%;  Pred. No. 4e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1  VPGG 4
              ||||
Db        27  VPGG 30


RESULT 15
Q53455
ID   Q53455          PRELIMINARY;      PRT;     61 AA.
AC   Q53455;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
```

```
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   ORF 5' OF AML (FRAGMENT).
OS   Streptomyces lividans.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1916;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94297727; PubMed=8025672;
RA   Virolle M.J., Gagnat J.;
RT   "Sequences involved in growth-phase-dependent expression and glucose
RT   repression of a Streptomyces alpha-amylase gene.";
RL   Microbiology 140:1059-1067(1994).
DR   EMBL; S73706; AAB31287.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   61 AA;   6569 MW;   655F08B04E525322 CRC64;

Query Match             100.0%;  Score 23;  DB 2;  Length 61;
Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1  VPGG 4
              ||||
Db        58  VPGG 61


Search completed: April 24, 2001, 16:40:39
Job time: 429 sec
```

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2001, 16:38:20 ; Search time 115.25 seconds
(without alignments)
14.880 Million cell updates/sec

Title: US-09-340-736-8
Perfect score: 174
Sequence: 1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description			
1	109	62.6	123	17	R95138	Silk like protein			
2	109	62.6	123	22	B64033	SELP2-SLPF functio			
3	109	62.6	159	17	R95139	Silk like protein			
4	109	62.6	159	22	B64034	SELP3-SLPF functio			
5	96	55.2	131	21	B19198	Amino acid sequenc			
6	96	55.2	136	21	B19196	Amino acid sequenc			
7	95.5	54.9	294	9	P82484	Tropoelastin. Gal			
8	91	52.3	126	21	B19197	Amino acid sequenc			
9	91	52.3	521	18	W36054	Mouse occludin pro			
10	91	52.3	646	18	W27178	Nephila clavipes s			
11	88.5	50.9	93	21	G24552	Arabidopsis thalia			

12	88.5	50.9	104	21	G24551	Arabidopsis thalia
13	88.5	50.9	133	21	G30160	Arabidopsis thalia
14	88.5	50.9	158	21	G52853	Arabidopsis thalia
15	88.5	50.9	158	21	G52859	Arabidopsis thalia
16	88.5	50.9	181	21	G52852	Arabidopsis thalia
17	88.5	50.9	181	21	G30158	Arabidopsis thalia
18	88.5	50.9	181	21	G52851	Arabidopsis thalia
19	88.5	50.9	263	21	G36620	Arabidopsis thalia
20	88.5	50.9	273	21	G36619	Arabidopsis thalia
21	88.5	50.9	309	21	G36618	Arabidopsis thalia
22	88	50.6	604	16	R99057	Spider dragline va
23	88	50.6	606	16	R99055	Spider dragline va
24	88	50.6	606	20	Y40101	Polymer of an anal
25	88	50.6	606	20	Y40102	Polymer of an anal
26	88	50.6	651	20	Y40097	Spider silk protei
27	88	50.6	718	12	R14308	N.clavipes draglin
28	88	50.6	718	19	W53346	Spider silk protei
29	88	50.3	718	21	Y59070	N. clavipes spider
30	87.5	49.7	132	21	B19199	Amino acid sequenc
31	86.5	49.7	399	21	G17435	Arabidopsis thalia
32	86.5	49.7	399	21	G20657	Arabidopsis thalia
33	86.5	49.7	421	21	G17434	Arabidopsis thalia
34	86.5	49.7	421	21	G20656	Arabidopsis thalia
35	86.5	49.7	612	21	G17433	Arabidopsis thalia
36	86.5	49.7	612	21	G20655	Arabidopsis thalia
37	85	48.9	176	21	G11337	Arabidopsis thalia
38	85	48.9	187	21	G11336	Arabidopsis thalia
39	85	48.9	206	22	G11335	Arabidopsis thalia
40	85	48.9	235	17	N05704	Glycine-rich repea
41	85	48.9	235	19	W79126	Epstein Barr Virus
42	85	48.9	261	19	W79137	FIGA Gly-ala inser
43	85	48.9	641	20	Y28843	Epstein Barr Virus
44	85	48.9	641	21	Y95856	Epstein Barr Virus
45	84.5	48.6	64	21	G16184	Arabidopsis thalia

ALIGNMENTS

RESULT 1
R95138
ID R95138 standard; Protein; 123 AA.
XX
AC R95138;
XX
DT 03-FEB-1997 (first entry)
XX
DE Silk like protein (SELP)2-SLPF.
XX
KW Polymer; repeat unit; natural fibroin; intervening oligopeptide;
KW fibre; film; membrane; emulsion; coating; silk like protein;
KW specific binding material; catalyst; purification agent; composite;
KW laminate; adhesive; cell growth surface; affinity column;
KW biological material support; wound dressing; in vivo prothesis.
XX
OS Synthetic.
XX
PN US5514581-A.
XX
PD 07-MAY-1996.
XX
PF 04-NOV-1986; 86US-0927258.
XX
PR 06-NOV-1990; 90US-0609716.
PR 04-NOV-1986; 86US-0927258.
PR 29-OCT-1987; 87US-0114618.
PR 09-NOV-1988; 88US-0269429.
PR 07-NOV-1989; 89WO-US05016.
XX
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI Cappello J, Ferrari FA;
XX

DR   WPI; 1996-238772/24.
XX
PT   DNA encoding protein contg. repeated fibroin derived segments - linked by oligopeptide with cell adhesion properties useful, e.g. in wound dressings
XX
PS   Example 3; Columns 127-128; 71pp; English.
XX
CC   Novel DNA sequence encodes a polymer comprising segments of repeating units of 3-9 amino acids from natural fibroin, i.e. the present silk like protein (SELP)2-SLPF sequence, able to assemble into aligned structures formable into articles. The polymer comprises at least 2 segments joined by an unaligned intervening oligopeptide, other than the repeating unit. The polymer can be used to make fibres, films, membranes, emulsions, coatings, etc., useful as, e.g. specific binding materials, catalysts, purificn. agents, laminates, adhesives, cell growth surfaces, affinity columns and supports for biological materials. Typical applications include wound dressings, and in vivo protheses. The polymer produces articles with good mechanical properties, and the intervening oligopeptide can provide a ligand for binding a mol., antibody, etc., or a chemically reactive site for coupling to proteins, etc..
XX
SQ   Sequence 123 AA;

Query Match   62.6%; Score 109; DB 17; Length 123;
Best Local Similarity 62.1%; Pred. No. 4.5e-06;
Matches 18; Conservative   6; Mismatches   5; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
       ||:| ||:| ||:| ||:| ||:| ||:|
Db 35 ggvgvgvgvgvgvgvgvgvgvgvgvggvg 63

RESULT 2
B64033
ID   B64033 standard; Protein; 123 AA.
XX
AC   B64033;
XX
DT   19-MAR-2001 (first entry)
XX
DE   SELP2-SLPF functional polymer amino acid sequence SEQ ID 103.
XX
KW   Proteinaceous polymer; repeat unit; structural polymer; coating; film; fibre; membrane; adhesive; emulsion; laminate; keratin; collagen.
XX
OS   Synthetic.
XX
OS   US6140072-A.
XX
PD   31-OCT-2000.
XX
PF   07-JUN-1995; 95US-0475411.
XX
PR   06-NOV-1990; 90US-0609716.
PR   04-NOV-1986; 86US-0927258.
PR   29-OCT-1987; 87US-0114618.
PR   09-NOV-1988; 88US-0269429.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J, Ferrari FA;
XX
DR   WPI; 2001-048958/06.
XX
PT   New DNA encoding a polymer with strands of repeating units of natural protein joined by intervening oligopeptide for producing high molecular weight polymers of amino acids -
XX
PS   Example 2; Column 49-50; 73pp; English.

XX
CC   This invention relates to DNA encoding a proteinaceous polymer. The polymer comprises strands of repeating units of a natural protein capable of assembling into aligned structures, with at least 2 strands joined by an intervening oligopeptide other than the repeating units. The intervening oligopeptide is unaligned and the polymer has individual strands of the same or different repeating units. The DNA is useful for producing high molecular weight polymers of amino acids based on biologically and chemically active structural polymers. These polymers may be used to provide a variety of structures for different purposes, and to produce articles including coatings, or other (nonstructural) components, e.g. fibres, films, membranes, adhesives or emulsions, or with other compounds to form composites or laminates. Peptide sequences B63977-B63991 represent monomer sequences which can be used in the polymers of the invention. Oligonucleotide sequences F23370 - F23386 and amino acid sequences B63992 - B64002 are used in the construction of SLP and FCB-SLP polymers. Oligonucleotide sequences F23387 - F23397 and amino acid sequences B64003 - B64008 are used in the construction of CLP (collagen like protein) polymers. Oligonucleotide sequences F23398 - F23409 and amino acid sequences B64009 - B64014 are used in the construction of keratin polymers. Proteins and peptides represented by sequences B64015 - B64049 are examples of polymers of the invention.
XX
SQ   Sequence 123 AA;

Query Match   62.6%; Score 109; DB 22; Length 123;
Best Local Similarity 62.1%; Pred. No. 4.5e-06;
Matches 18; Conservative   6; Mismatches   5; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
       ||:| ||:| ||:| ||:| ||:|
Db 35 ggvgvgvgvgvgvgvgvgvgvgvgvggvg 63

RESULT 3
R95139
ID   R95139 standard; Protein; 159 AA.
XX
AC   R95139;
XX
DT   03-FEB-1997 (first entry)
XX
DE   Silk like protein (SELP)3-SLPF.
XX
KW   Polymer; repeat unit; natural fibroin; intervening oligopeptide; fibre; film; membrane; emulsion; coating; silk like protein; specific binding material; catalyst; purification agent; composite; laminate; adhesive; cell growth surface; affinity column; biological material support; wound dressing; in vivo prothesis.
XX
OS   Synthetic.
XX
PN   US5514581-A.
XX
PD   07-MAY-1996.
XX
PF   04-NOV-1986; 86US-0927258.
XX
PR   06-NOV-1990; 90US-0609716.
PR   04-NOV-1986; 86US-0927258.
PR   29-OCT-1987; 87US-0114618.
PR   09-NOV-1988; 88US-0269429.
PR   09-NOV-1989; 89WO-US05016.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J, Ferrari FA;
XX
DR   WPI; 1996-238772/24.
XX
PT   DNA encoding protein contg. repeated fibroin derived segments -

PT   linked by oligopeptide with cell adhesion properties useful, e.g. in
PT   wound dressings
XX
PS   Example 3: Columns 129-130; 71pp; English.
XX
CC   Novel DNA sequence encodes a polymer comprising segments of
CC   repeating units of 3-9 amino acids from natural fibroin, i.e. the
CC   present silk like protein (SELP)3-SLPF sequence, able to assemble
CC   into aligned structures formable into articles. The polymer
CC   comprises at least 2 segments joined by an unaligned intervening
CC   oligopeptide, other than the repeating unit.
CC   The polymer can be used to make fibres, films, membranes,
CC   emulsions, coatings, etc., useful as, e.g. specific binding
CC   materials, catalysts, purificn. agents, composites, laminates,
CC   adhesives, cell growth surfaces, affinity columns and supports for
CC   biological materials. Typical applications include wound
CC   dressings, and in vivo prostheses. The polymer produces articles
CC   with good mechanical properties, and the intervening oligopeptide
CC   can provide a ligand for binding a mol., antibody, etc., or a
CC   chemically reactive site for coupling to proteins, etc..
XX
SQ   Sequence 159 AA;

Query Match   62.5%; Score 109; DB 17; Length 159;
Best Local Similarity   62.1%; Pred. No. 5.7e-06;
Matches   18; Conservative   6; Mismatches   5; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
              |||:||| |||:||| |||:||| |||:|
Db   53 ggvgvgvgvgvgvgvgvgvgvgvgvgvgvg 81

RESULT 4
B64034
ID   B64034 standard; Protein; 159 AA.
XX
AC   B64034;
XX
DT   19-MAR-2001 (first entry)
XX
DE   SELP3-SLPF functional polymer amino acid sequence SEQ ID 104.
XX
KW   Proteinaceous polymer; repeat unit; structural polymer; coating; film;
KW   fibre; membrane; adhesive; emulsion; laminate; keratin; collagen.
XX
OS   Synthetic.
XX
PN   US6140072-A.
XX
PD   31-OCT-2000.
XX
PF   07-JUN-1995; 95US-0475411.
XX
PR   06-NOV-1990; 90US-0609716.
PR   04-NOV-1986; 86US-0927258.
PR   29-OCT-1987; 87US-0114618.
PR   09-NOV-1988; 88US-0269429.
XX
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI   Cappello J, Ferrari FA;
XX
DR   WPI; 2001-048958/06.
XX
PT   New DNA encoding a polymer with strands of repeating units of natural
PT   protein joined by intervening oligopeptide for producing high molecular
PT   weight polymers of amino acids
XX
PS   Example 2; Column 49-50; 73pp; English.
XX
CC   This invention relates to DNA encoding a proteinaceous polymer. The
CC   polymer comprises strands of repeating units of a natural protein capable

CC   of assembling into aligned structures, with at least 2 strands joined by
CC   an intervening oligopeptide other than the repeating units. The
CC   intervening oligopeptide is unaligned and the polymer has individual
CC   strands of the same or different repeating units. The DNA is useful for
CC   producing high molecular weight polymers of amino acids based on
CC   biologically and chemically active structural polymers. These polymers
CC   may be used to provide a variety of structures for different purposes,
CC   and to produce articles including coatings, or other (non)structural
CC   components, e.g. fibres, films, membranes, adhesives or emulsions, or
CC   with other compounds and/or compositions to form composites or laminates.
CC   Peptide sequences B63971-B63991 represent monomer sequences which can be
CC   used in the polymers of the invention. Oligonucleotide sequences
CC   F23370 - F23386 and amino acid sequences B63992 - B64002 are used in the
CC   construction of SLP and FCB-SLP polymers. Oligonucleotide sequences
CC   F23387 - F23397 and amino acid sequences B64003 - B64008 are used in the
CC   construction of CLP (collagen like protein) polymers. Oligonucleotide
CC   sequences F23398 - F23409 and amino acid sequences B64009 - B64014 are
CC   used in the construction of keratin polymers. Proteins and peptides
CC   represented by sequences B64015 - B64049 are examples of polymers of the
CC   invention.
XX
SQ   Sequence 159 AA;

Query Match   62.6%; Score 109; DB 22; Length 159;
Best Local Similarity   62.1%; Pred. No. 5.7e-06;
Matches   18; Conservative   6; Mismatches   5; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
              |||:||| |||:||| |||:||| |||:|
Db   53 ggvgvgvgvgvgvgvgvgvgvgvgvgvgvg 81

RESULT 5
B19198
ID   B19198 standard; Protein; 131 AA.
XX
AC   B19198;
XX
DT   19-FEB-2001 (first entry)
XX
DE   Amino acid sequence of a bay scallop abduction polypeptide.
XX
KW   Bay scallop; abductin; chemomechanical transduction; drug delivery;
KW   inverse temperature transition; water soluble drug; bionaterial;
KW   fabric; organ prosthesis.
XX
OS   Argopecten sp.
XX
PN   US6127166-A.
XX
PD   03-OCT-2000.
XX
PF   03-NOV-1997; 97US-0963168.
PR   03-NOV-1997; 97US-0963168.
XX
PA   (BAYL/) BAYLEY H.
PA   (CACQ/) CAO Q.
PA   (WANG/) WANG Y.
XX
PI   Bayley H, Cao Q, Wang Y;
XX
DR   WPI; 2000-611057/58.
DR   N-PSDB; C613379.
XX
PT   Abductin nucleic acid molecules, useful for expressing abductin
PT   polypeptides which are used in the manufacture of drug delivery
PT   vehicles for administering water soluble drugs
XX
PS   Example 1; Fig 2; 30pp; English.
XX
CC   The present sequence represents a bay scallop abductin polypeptide.

```
CC   The polypeptide is capable of chemomechanical transduction or inverse
CC   temperature transition. Abductin polypeptides contain glycine-rich
CC   sequences. Abductin polypeptides are useful in the manufacture of
CC   drug delivery vehicles for administering water soluble drugs. The
CC   abductin polypeptides and their derivatives are also useful in the
CC   manufacture of broad range of biomaterials ranging from light-weight
CC   durable fabric for clothing to matrices useful for human tissue and
CC   organ prostheses.
XX
SQ   Sequence   131 AA;

Query Match            55.2%;  Score 96;  DB 21;  Length 131;
Best Local Similarity  50.0%;  Pred. No. 0.00014;
Matches   17;  Conservative    6;  Mismatches   7;  Indels    4;  Gaps    1;

Qy    1 GGLGYGGLGYGGLGYGGL---GYGGLGY 30
         ||:|| |:||:|| |:||:  |:||:||
Db   98 ggmgggnagfggmgggnagfggmgggggfggkky 131

RESULT 6
ID   B19196 standard; Protein; 136 AA.
XX
AC   B19196;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Amino acid sequence of a bay scallop abduction polypeptide.
XX
KW   Bay scallop; abductin; chemomechanical transduction; drug delivery;
KW   inverse temperature transition; water soluble drug; biomaterial;
KW   fabric; organ prosthesis.
XX
OS   Argopecten sp.
XX
PN   US6127166-A.
XX
PD   03-OCT-2000.
XX
PF   03-NOV-1997;  97US-0963168.
XX
PR   03-NOV-1997;  97US-0963168.
XX
PA   (BAYL/) BAYLEY H.
PA   (CAOQ/) CAO Q.
PA   (WANG/) WANG Y.
XX
PI   Bayley H, Cao Q, Wang Y;
XX
     WPI; 2000-611057/58.
     N-PSDB; C61376, C61377.
XX
PT   Abductin nucleic acid molecules, useful for expressing abductin
PT   polypeptides which are used in the manufacture of drug delivery
PT   vehicles for administering water soluble drugs -
XX
PS   Example 1; Fig 2; 30pp; English.
XX
CC   The present sequence represents a bay scallop abductin polypeptide.
CC   The polypeptide is capable of chemomechanical transduction or inverse
CC   temperature transition. Abductin polypeptides contain glycine-rich
CC   sequences. Abductin polypeptides are useful in the manufacture of
CC   drug delivery vehicles for administering water soluble drugs. the
CC   abductin polypeptides and their derivatives are also useful in the
CC   manufacture of broad range of biomaterials ranging from light-weight
CC   durable fabric for clothing to matrices useful for human tissue and
CC   organ prostheses.
XX
SQ   Sequence   136 AA;


Query Match            55.2%;  Score 96;  DB 21;  Length 136;
Best Local Similarity  50.0%;  Pred. No. 0.00015;
Matches   17;  Conservative    6;  Mismatches   7;  Indels    4;  Gaps    1;

Qy    1 GGLGYGGLGYGGLGYGGL---GYGGLGY 30
         ||:|| |:||:|| |:||:  |:||:||
Db  103 cgmgggnagfggmgggnagfggmgggggfggkgy 136

RESULT   7
P82484
ID   P82484 standard; protein; 294 AA.
XX
AC   P82484;
XX
DT   31-OCT-1990  (first entry)
XX
DE   Tropoelastin.
XX
KW   Chicken; tropoelastin; platelet aggregation; arteriosclerosis.
XX
OS   Gallus domesticus.
XX
FH   Key              Location/Qualifiers
FT   Region           3..32
FT                    /label=repeat_region
FT                    /rpt_type=tandem
FT                    /rpt_unit= RU1:VPGVG
FT   Region           53..64
FT                    /label=repeat_unit
FT                    /rpt_type=direct
FT                    /rpt_unit= RU2:AAAAKAAAKAAK
FT   Region           92..96
FT                    /label=RU1
FT   Region           99..103
FT                    /label=RU1
FT   Region           119..123
FT                    /label=RU1
FT   Region           126..137
FT                    /label=RU2
FT   Region           173..182
FT                    /label=rpt_region
FT                    /rpt_type=tandem
FT                    /rpt_unit=RU1 (x2)
XX
PN   JP63014694-A.
XX
PD   21-JAN-1988.
XX
PF   08-JUL-1986;  86JP-0158655.
XX
PR   08-JUL-1986;  86JP-0158655.
XX
PA   (KAOS ) KAO CORP.
PA   (SHIS ) SHISEIDO KK.
XX
DR   WPI; 1988-059829/09.
DR   N-PSDB; N82138.
XX
PT   cDNA fragment encoding tropoelastin - has inhibiting activity towards
PT   platelet aggregation and is useful for prophylaxis of arteriosclerosis.
XX
PS   Claim 1; Fig 1; 5pp; Japanese.
XX
CC   The cDNA was sequenced from a clone prepd. from mRNA isolated from
CC   chicken aorta tissue. It can be used to produce tropoelastin which
CC   is a precursor of elastin which does not have desmosine, isodesmosine
CC   and lysinoylleucine residues and isn't cross-linked.
XX
SQ   Sequence   294 AA;

Query Match            54.9%;  Score 95.5;  DB 9;  Length 294;
```

```
Best Local Similarity   69.0%;  Pred. No. 0.00033;
Matches  20;  Conservative  1;  Mismatches  7;  Indels  1;  Gaps  1;

Qy    1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
          ||| |||| |||| |||| ||| |:| |
Db  218 gglgyvgglgvvgglgagglg-agvqvpgfg 245

RESULT 8
ID   B19197 standard; Protein; 126 AA.
XX
AC   B19197;
XX
DT   19-FEB-2001 (first entry)
XX
DE   Amino acid sequence of a bay scallop abduction polypeptide.

KW   Bay scallop; abductin; chemomechanical transduction; drug delivery;
     inverse temperature transition; water soluble drug; biomaterial;
KW   fabric; organ prosthesis.
XX
OS   Argopecten sp.
XX
PN   US6127166-A.
XX
PD   03-OCT-2000.
XX
PF   03-NOV-1997;  97US-0963168.
PR   03-NOV-1997;  97US-0963168.
XX
PA   (BAYL/) BAYLEY H.
PA   (CAOQ/) CAO Q.
PA   (WANG/) WANG Y.
PI   Bayley H. Cao Q. Wang Y;
XX
DR   WPI; 2000-611057/58.
DR   N-PSDB: C61378.
XX
PT   Abductin nucleic acid molecules, useful for expressing abductin
PT   polypeptides which are used in the manufacture of drug delivery
PT   vehicles for administering water soluble drugs -
XX
     Example 1; Fig 2; 30pp; English.
XX
CC   The present sequence represents a bay scallop abductin polypeptide.
CC   The polypeptide is capable of chemomechanical transduction or inverse
CC   temperature transition. Abductin polypeptides contain glycine-rich
CC   sequences. Abductin polypeptides are useful in the manufacture of
CC   drug delivery vehicles for administering water soluble drugs. The
CC   abductin polypeptides and their derivatives are also useful in the
CC   manufacture of broad range of biomaterials ranging from light-weight
CC   durable fabric for clothing to matrices useful for human tissue and
CC   organ prostheses.
XX
SQ   Sequence  126 AA;

Query Match          52.3%;  Score 91;  DB 21;  Length 126;
Best Local Similarity 50.0%;  Pred. No. 0.0005;
Matches  17;  Conservative  6;  Mismatches  7;  Indels  4;  Gaps  1;

Qy    1 GGLGYGGLGYGGLGYGGLGYGGL----GYGGLGY 30
          |:||:| |:||:| |:||:|     ||||||
Db   93 ggmgsmggifggmggcgnagfggmggcggfgkgy 126

RESULT 9
ID   W36054 standard; Protein; 521 AA.
XX
AC   W36054;
XX
DT   25-MAR-1998 (first entry)
XX
DE   Mouse occludin protein sequence.
XX
KW   Occludin; adhesion molecule; membrane tight junction;
KW   occludin localisation; membrane; occludin expression;
KW   blood-brain barrier disorder.
XX
OS   Mus sp.
XX
PN   WO9732982-A1.
XX
PD   12-SEP-1997.
XX
PF   05-MAR-1997;  97WO-JP00665.
XX
PR   12-DEC-1996;  96JP-0331944.
PR   07-MAR-1996;  96JP-0049880.
XX
PA   (EISA ) EISAI CO LTD.
XX
PI   Tsukita S;
XX
DR   WPI; 1997-470546/43.
DR   N-PSDB: T097974.
XX
PT   DNA encoding human, dog and mouse occludin(s) - useful for screening
PT   for substances influencing occluding expression
XX
PS   Claim 5; Pages 20-22; 36pp; Japanese.
XX
CC   The present sequence represents mouse occludin, a novel adhesion
CC   molecule which is a constituent protein of membrane tight junctions.
CC   Antibodies raised against the protein are used to assay for occludin
CC   in samples, and for the study of occludin localisation in membranes
CC   (e.g. by immunofluorescent cell-staining). The transformants and
CC   antibodies are used for screening of substances which potentially
CC   influence occludin expression. Therapeutic polynucleotides derived
CC   from the DNA encoding the present sequence are used for treatment of
CC   disorders involving the blood-brain barrier. Primers that hybridise to
CC   the occludin DNA sequence can be used to detect occludin DNA by PCR.
XX
SQ   Sequence  521 AA;

Query Match          52.3%;  Score 91;  DB 18;  Length 521;
Best Local Similarity 73.1%;  Pred. No. 0.0018;
Matches  19;  Conservative  0;  Mismatches  5;  Indels  2;  Gaps  2;

Qy    5 YGGLGYGGLGYGGLGYGGLGYGGLGY 30
          | |||| |||| |||| ||| ||
Db  106 ysgfgysg-gygg-gysgygysyggy 129

RESULT 10
ID   W27178 standard; Protein; 646 AA.
XX
AC   W27178;
XX
DT   09-DEC-1997 (first entry)
XX
DE   Nephila clavipes spider silk protein.
XX
KW   High strength film; fibre; woven article; parachutes; sails;
KW   absorber; body armour; heavy metal; biological weapon; chemical;
KW   flavour; fragrance; Nephila clavipes.
XX
OS   Nephila clavipes.
XX
PN   WO9708315-A1.
```

```
XX
PD   06-MAR-1997.
XX
PF   22-AUG-1996;  96WO-US13767.
XX
PR   22-AUG-1995;  95US-0517694.
XX
PA   (BASE/) BASEL R M.
PA   (ELIO/) ELION G R.
XX
PI   Basel RM, Elion GR;
XX
DR   WPI; 1997-179272/16.
DR   N-PSDB; T85356.
XX
PT   New opt. multimerised DNA sequences encoding spider silk protein -
PT   contg. both repetitive and non-repetitive sequences, useful for
PT   making high strength films, fibres, woven articles etc.
XX
PS   Example 2; Fig 1; 57pp; English.
XX
CC   A process has been developed for the production of a DNA fragment
CC   encoding silk protein. The process involves: (a) selecting target DNA,
CC   from a silk-producing spider, that contains many repetitive and non-
CC   repetitive regions; (b) selecting a single-stranded DNA primer of at
CC   least 10 nucleotides with a sequence that is complementary to a region
CC   of the target; (c) repetitively combining the primer with melted target
CC   DNA, incubating the mixture with nucleotides and a DNA polymerase with
CC   proofreading activity to produce a DNA fragment which is complementary
CC   to the target and is at least 2 kb long. The present sequence encodes
CC   the spider silk protein from Nephila clavipes. The DNA fragment can be
CC   used to make fibres, films, woven articles, e.g. for use in parachutes,
CC   sails, body armour, and absorbers (e.g. of heavy metals, biological
CC   weapons, DNA, chemicals, flavours and fragrances). The high molecular
CC   weight (90-250 kD) of spider silk proteins can be produced on a
CC   commercial scale (at over 2 g/l cell mass). It has better tensile
CC   strength and elasticity than silkworm silk. Inclusion of both repetitive
CC   and non-repetitive regions ensures isolation of stable clones
XX
SQ   Sequence 646 AA;

Query Match             51.1%;  Score 89;  DB 18;  Length 646;
Best Local Similarity   69.2%;  Pred. No. 0.0037;
Matches   18; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

Qy     4 GYGGLGYGGLGYGGLGYGGLGYGGLG 29
           |||| | |  |||||| | | |||
Db   133 gygggasgagqsgyrgglgsqgagrgglg 158
```

```
IT 11
(552
ID   G24552 standard;  Protein; 93 AA.
XX
AC   G24552;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 28269.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
```

```
PR   25-FEB-1999;    99US-0121825.
PR   05-MAR-1999;    99US-0123180.
PR   09-MAR-1999;    99US-0123548.
PR   23-MAR-1999;    99US-0125788.
PR   25-MAR-1999;    99US-0125764.
PR   29-MAR-1999;    99US-0125785.
PR   01-APR-1999;    99US-0127462.
PR   06-APR-1999;    99US-0128234.
PR   08-APR-1999;    99US-0128714.
PR   16-APR-1999;    99US-0129845.
PR   19-APR-1999;    99US-0130077.
PR   21-APR-1999;    99US-0130449.
PR   23-APR-1999;    99US-0130510.
PR   23-APR-1999;    99US-0130891.
PR   28-APR-1999;    99US-0131449.
PR   30-APR-1999;    99US-0132048.
PR   30-APR-1999;    99US-0132207.
PR   04-MAY-1999;    99US-0132484.
PR   05-MAY-1999;    99US-0132485.
PR   06-MAY-1999;    99US-0132486.
PR   06-MAY-1999;    99US-0132487.
PR   07-MAY-1999;    99US-0132863.
PR   11-MAY-1999;    99US-0134256.
PR   14-MAY-1999;    99US-0134218.
PR   14-MAY-1999;    99US-0134219.
PR   14-MAY-1999;    99US-0134221.
PR   14-MAY-1999;    99US-0134370.
PR   18-MAY-1999;    99US-0134768.
PR   19-MAY-1999;    99US-0134941.
PR   20-MAY-1999;    99US-0135124.
PR   21-MAY-1999;    99US-0135353.
PR   24-MAY-1999;    99US-0135629.
PR   25-MAY-1999;    99US-0136021.
PR   27-MAY-1999;    99US-0136392.
PR   28-MAY-1999;    99US-0136782.
PR   01-JUN-1999;    99US-0137222.
PR   03-JUN-1999;    99US-0137528.
PR   04-JUN-1999;    99US-0137502.
PR   07-JUN-1999;    99US-0137724.
PR   08-JUN-1999;    99US-0138094.
PR   10-JUN-1999;    99US-0138540.
PR   10-JUN-1999;    99US-0138847.
PR   14-JUN-1999;    99US-0139119.
PR   16-JUN-1999;    99US-0139452.
PR   16-JUN-1999;    99US-0139453.
PR   17-JUN-1999;    99US-0139492.
PR   18-JUN-1999;    99US-0139454.
PR   18-JUN-1999;    99US-0139455.
PR   18-JUN-1999;    99US-0139456.
PR   18-JUN-1999;    99US-0139457.
PR   18-JUN-1999;    99US-0139458.
PR   18-JUN-1999;    99US-0139459.
PR   18-JUN-1999;    99US-0139460.
PR   18-JUN-1999;    99US-0139461.
PR   18-JUN-1999;    99US-0139462.
PR   18-JUN-1999;    99US-0139463.
PR   18-JUN-1999;    99US-0139750.
PR   18-JUN-1999;    99US-0139763.
PR   21-JUN-1999;    99US-0139817.
PR   22-JUN-1999;    99US-0139899.
PR   23-JUN-1999;    99US-0140353.
PR   23-JUN-1999;    99US-0140354.
PR   24-JUN-1999;    99US-0140695.
PR   28-JUN-1999;    99US-0140823.
PR   29-JUN-1999;    99US-0140991.
PR   30-JUN-1999;    99US-0141287.
PR   01-JUL-1999;    99US-0141842.
PR   01-JUL-1999;    99US-0142154.
PR   02-JUL-1999;    99US-0142055.
PR   06-JUL-1999;    99US-0142390.
PR   08-JUL-1999;    99US-0142803.
PR   09-JUL-1999;    99US-0142920.
PR   12-JUL-1999;    99US-0142977.
```

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145292.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145275.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149275.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150056.
PR 27-AUG-1999; 99US-0150084.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match   50.9%;   Score 88.5; DB 21; Length 93;
Best Local Similarity   55.5%;   Pred. No. 0.00073;
Matches 20; Conservative   1; Mismatches   8;   Indels   7;   Gaps   1;

Qy   1 GGLGYGGLGYG---------GLGYGGLGYGGLGYGGLG 29
           || ||| |||| ||| ||| |:|| |||
Db   55 ggaglgglgtgagicagaglglgagggdfgggagggglg 90

RESULT 12
G24551
ID   G24551 standard; Protein; 104 AA.
XX
AC   G24551;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 28268.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.

```
PR  06-APR-1999;  99US-0128234.        PR  19-JUL-1999;  99US-0144332.
PR  08-APR-1999;  99US-0128714.        PR  19-JUL-1999;  99US-0144333.
PR  16-APR-1999;  99US-0129845.        PR  19-JUL-1999;  99US-0144334.
PR  19-APR-1999;  99US-0130077.        PR  19-JUL-1999;  99US-0144335.
PR  21-APR-1999;  99US-0130449.        PR  20-JUL-1999;  99US-0144352.
PR  23-APR-1999;  99US-0130510.        PR  20-JUL-1999;  99US-0144632.
PR  23-APR-1999;  99US-0130891.        PR  20-JUL-1999;  99US-0144884.
PR  28-APR-1999;  99US-0131449.        PR  21-JUL-1999;  99US-0144814.
PR  30-APR-1999;  99US-0132048.        PR  21-JUL-1999;  99US-0145086.
PR  30-APR-1999;  99US-0132407.        PR  22-JUL-1999;  99US-0145088.
PR  04-MAY-1999;  99US-0132484.        PR  22-JUL-1999;  99US-0145085.
PR  05-MAY-1999;  99US-0132485.        PR  22-JUL-1999;  99US-0145087.
PR  06-MAY-1999;  99US-0132486.        PR  22-JUL-1999;  99US-0145089.
PR  06-MAY-1999;  99US-0132487.        PR  23-JUL-1999;  99US-0145192.
PR  07-MAY-1999;  99US-0132863.        PR  23-JUL-1999;  99US-0145145.
PR  11-MAY-1999;  99US-0134256.        PR  23-JUL-1999;  99US-0145218.
PR  14-MAY-1999;  99US-0134218.        PR  26-JUL-1999;  99US-0145224.
PR  14-MAY-1999;  99US-0134219.        PR  27-JUL-1999;  99US-0145276.
PR  14-MAY-1999;  99US-0134221.        PR  27-JUL-1999;  99US-0145913.
PR  18-MAY-1999;  99US-0134370.        PR  27-JUL-1999;  99US-0145918.
PR  19-MAY-1999;  99US-0134768.        PR  28-JUL-1999;  99US-0145919.
PR  20-MAY-1999;  99US-0134941.        PR  02-AUG-1999;  99US-0145951.
PR  21-MAY-1999;  99US-0135124.        PR  02-AUG-1999;  99US-0146386.
PR  24-MAY-1999;  99US-0135353.        PR  02-AUG-1999;  99US-0146388.
PR  25-MAY-1999;  99US-0135629.        PR  03-AUG-1999;  99US-0146389.
PR  27-MAY-1999;  99US-0136021.        PR  04-AUG-1999;  99US-0147038.
PR  28-MAY-1999;  99US-0136392.        PR  04-AUG-1999;  99US-0147204.
PR  01-JUN-1999;  99US-0136782.        PR  05-AUG-1999;  99US-0147302.
PR  03-JUN-1999;  99US-0137222.        PR  06-AUG-1999;  99US-0147192.
PR  04-JUN-1999;  99US-0137528.        PR  06-AUG-1999;  99US-0147260.
PR  07-JUN-1999;  99US-0137502.        PR  09-AUG-1999;  99US-0147303.
PR  08-JUN-1999;  99US-0137724.        PR  09-AUG-1999;  99US-0147416.
PR  10-JUN-1999;  99US-0138094.        PR  10-AUG-1999;  99US-0147493.
PR  14-JUN-1999;  99US-0138540.        PR  11-AUG-1999;  99US-0147935.
PR  16-JUN-1999;  99US-0138847.        PR  12-AUG-1999;  99US-0148171.
PR  16-JUN-1999;  99US-0139119.        PR  13-AUG-1999;  99US-0148319.
PR  17-JUN-1999;  99US-0139452.        PR  15-AUG-1999;  99US-0148341.
PR  18-JUN-1999;  99US-0139453.        PR  17-AUG-1999;  99US-0148565.
PR  18-JUN-1999;  99US-0139492.        PR  18-AUG-1999;  99US-0148684.
PR  18-JUN-1999;  99US-0139454.        PR  20-AUG-1999;  99US-0149368.
PR  18-JUN-1999;  99US-0139455.        PR  20-AUG-1999;  99US-0149175.
PR  18-JUN-1999;  99US-0139456.        PR  20-AUG-1999;  99US-0149426.
PR  18-JUN-1999;  99US-0139457.        PR  23-AUG-1999;  99US-0149722.
PR  18-JUN-1999;  99US-0139458.        PR  25-AUG-1999;  99US-0149723.
PR  18-JUN-1999;  99US-0139459.        PR  26-AUG-1999;  99US-0149929.
PR  18-JUN-1999;  99US-0139460.        PR  27-AUG-1999;  99US-0149902.
PR  18-JUN-1999;  99US-0139461.        PR  27-AUG-1999;  99US-0149930.
PR  18-JUN-1999;  99US-0139462.        PR  30-AUG-1999;  99US-0150566.
PR  18-JUN-1999;  99US-0139463.        PR  01-SEP-1999;  99US-0150884.
PR  21-JUN-1999;  99US-0139750.        PR  07-SEP-1999;  99US-0151065.
PR  23-JUN-1999;  99US-0139817.        PR  10-SEP-1999;  99US-0151066.
PR  23-JUN-1999;  99US-0139899.        PR  13-SEP-1999;  99US-0151080.
PR  24-JUN-1999;  99US-0140353.        PR  15-SEP-1999;  99US-0151303.
PR  24-JUN-1999;  99US-0140354.        PR  16-SEP-1999;  99US-0151438.
PR  28-JUN-1999;  99US-0140695.        PR  20-SEP-1999;  99US-0151930.
PR  29-JUN-1999;  99US-0140823.        PR  23-SEP-1999;  99US-0152363.
PR  30-JUN-1999;  99US-0140991.        PR  24-SEP-1999;  99US-0153070.
PR  01-JUL-1999;  99US-0141287.        PR  28-SEP-1999;  99US-0153758.
PR  01-JUL-1999;  99US-0141842.        PR  29-SEP-1999;  99US-0154018.
PR  02-JUL-1999;  99US-0142154.        PR  04-OCT-1999;  99US-0154039.
PR  06-JUL-1999;  99US-0142055.        PR  05-OCT-1999;  99US-0154779.
PR  06-JUL-1999;  99US-0142390.        PR  06-OCT-1999;  99US-0155139.
PR  08-JUL-1999;  99US-0142803.        PR  07-OCT-1999;  99US-0155486.
PR  09-JUL-1999;  99US-0142920.        PR  08-OCT-1999;  99US-0155659.
PR  12-JUL-1999;  99US-0142977.        PR  12-OCT-1999;  99US-0156458.
PR  13-JUL-1999;  99US-0143624.        PR  13-OCT-1999;  99US-0156596.
PR  14-JUL-1999;  99US-0143542.                          99US-0157117.
PR  15-JUL-1999;  99US-0144005.                          99US-0157753.
PR  16-JUL-1999;  99US-0144085.                          99US-0157865.
PR  16-JUL-1999;  99US-0144086.                          99US-0158029.
PR  19-JUL-1999;  99US-0144325.                          99US-0158232.
PR  19-JUL-1999;  99US-0144331.                          99US-0158369.
                                                         99US-0159293.
```

PR	13-OCT-1999;	99US-0159294.	
PR	13-OCT-1999;	99US-0159295.	
PR	14-OCT-1999;	99US-0159329.	
PR	14-OCT-1999;	99US-0159330.	
PR	14-OCT-1999;	99US-0159331.	
PR	14-OCT-1999;	99US-0159637.	
PR	14-OCT-1999;	99US-0159638.	
PR	18-OCT-1999;	99US-0159584.	
PR	21-OCT-1999;	99US-0160741.	
PR	21-OCT-1999;	99US-0160758.	
PR	21-OCT-1999;	99US-0160767.	
PR	21-OCT-1999;	99US-0160770.	
PR	21-OCT-1999;	99US-0160814.	
PR	21-OCT-1999;	99US-0160815.	
PR	22-OCT-1999;	99US-0160980.	
PR	22-OCT-1999;	99US-0160981.	
PR	25-OCT-1999;	99US-0160989.	
PR	25-OCT-1999;	99US-0161404.	
PR	25-OCT-1999;	99US-0161405.	
PR	26-OCT-1999;	99US-0161406.	
PR	26-OCT-1999;	99US-0161359.	
PR	26-OCT-1999;	99US-0161360.	
PR	26-OCT-1999;	99US-0161361.	
PR	28-OCT-1999;	99US-0161920.	
PR	28-OCT-1999;	99US-0161992.	
PR	28-OCT-1999;	99US-0161993.	
PR	29-OCT-1999;	99US-0162142.	

```
Query Match        50.9%;   Score 88.5;  DB 21;  Length 104;
Best Local Similarity  55.6%;  Pred. No. 0.00081;
Matches  20;  Conservative  1;  Mismatches  8;  Indels  7;  Gaps  1;

Qy    1 GGLGYGGLGYG--------GLGYGGLGYGGLGYGGLG  29
          ||  ||  |||||     ||     ||| : ||   |||
Db   66 ggaglggligyagigagagigigsggfggggaggig 101
```

RESULT 13
G30160
ID   G30160 standard; Protein; 133 AA.
XX
AC   G30160;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 36008.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX

PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	

PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134255.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134222.	
PR	18-MAY-1999;	99US-0134770.	
PR	19-MAY-1999;	99US-0134768.	
PR	20-MAY-1999;	99US-0134941.	
PR	21-MAY-1999;	99US-0135124.	
PR	24-MAY-1999;	99US-0135353.	
PR	25-MAY-1999;	99US-0135629.	
PR	27-MAY-1999;	99US-0136021.	
PR	28-MAY-1999;	99US-0136392.	
PR	01-JUN-1999;	99US-0136782.	
PR	03-JUN-1999;	99US-0137222.	
PR	04-JUN-1999;	99US-0137528.	
PR	07-JUN-1999;	99US-0137502.	
PR	08-JUN-1999;	99US-0137724.	
PR	10-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	14-JUN-1999;	99US-0138847.	
PR	16-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	
PR	17-JUN-1999;	99US-0139453.	
PR	18-JUN-1999;	99US-0139454.	
PR	18-JUN-1999;	99US-0139455.	
PR	18-JUN-1999;	99US-0139456.	
PR	18-JUN-1999;	99US-0139457.	
PR	18-JUN-1999;	99US-0139458.	
PR	18-JUN-1999;	99US-0139459.	
PR	18-JUN-1999;	99US-0139460.	
PR	18-JUN-1999;	99US-0139461.	
PR	18-JUN-1999;	99US-0139462.	
PR	18-JUN-1999;	99US-0139463.	
PR	18-JUN-1999;	99US-0139750.	
PR	21-JUN-1999;	99US-0139763.	
PR	22-JUN-1999;	99US-0139817.	
PR	23-JUN-1999;	99US-0139899.	
PR	23-JUN-1999;	99US-0140354.	
PR	24-JUN-1999;	99US-0140695.	
PR	28-JUN-1999;	99US-0140823.	
PR	29-JUN-1999;	99US-0140991.	
PR	30-JUN-1999;	99US-0141287.	
PR	01-JUL-1999;	99US-0141842.	
PR	01-JUL-1999;	99US-0142154.	
PR	02-JUL-1999;	99US-0142055.	
PR	06-JUL-1999;	99US-0142390.	
PR	08-JUL-1999;	99US-0142803.	
PR	09-JUL-1999;	99US-0142920.	
PR	12-JUL-1999;	99US-0142977.	
PR	13-JUL-1999;	99US-0143542.	
PR	14-JUL-1999;	99US-0143624.	
PR	15-JUL-1999;	99US-0144005.	
PR	16-JUL-1999;	99US-0144085.	
PR	16-JUL-1999;	99US-0144086.	
PR	19-JUL-1999;	99US-0144335.	
PR	19-JUL-1999;	99US-0144331.	
PR	19-JUL-1999;	99US-0144332.	
PR	19-JUL-1999;	99US-0144333.	
PR	19-JUL-1999;	99US-0144334.	
PR	19-JUL-1999;	99US-0144335.	
PR	20-JUL-1999;	99US-0144352.	
PR	20-JUL-1999;	99US-0144632.	
PR	20-JUL-1999;	99US-0144884.	

PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   22-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   03-AUG-1999;   99US-0146389.
PR   04-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   05-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   06-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   09-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   10-AUG-1999;   99US-0147935.
PR   11-AUG-1999;   99US-0148171.
PR   12-AUG-1999;   99US-0148319.
PR   13-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   16-AUG-1999;   99US-0148684.
PR   17-AUG-1999;   99US-0149368.
PR   18-AUG-1999;   99US-0149175.
PR   20-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149723.
PR   23-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   25-AUG-1999;   99US-0149930.
PR   26-AUG-1999;   99US-0150566.
PR   27-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   30-AUG-1999;   99US-0151080.
PR   31-AUG-1999;   99US-0151103.
PR   01-SEP-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   14-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.

PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0150767.
PR   21-OCT-1999;   99US-0150768.
PR   21-OCT-1999;   99US-0150770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   25-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Query Match   50.9%; Score 88.5; DB 21; Length 133;
Best Local Similarity 66.7%; Pred. No. 0.001;
Matches   20; Conservative   1; Mismatches   8; Indels   1; Gaps   1;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
           || || ||| || ||| |||| || | ||
Db   105 gggygg ggggyygggggyygggdggggf 133

RESULT 14
G52853
ID   G52853 standard; Protein; 133 AA.
XX
AC   G52853;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 67230.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.

PR	Date	Number		PR	Date	Number
PR	07-MAY-1999;	99US-0132863.		PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134256.		PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.		PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.		PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.		PR	27-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	99US-0134370.		PR	27-JUL-1999;	99US-0145918.
PR	19-MAY-1999;	99US-0134768.		PR	27-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	99US-0134941.		PR	28-JUL-1999;	99US-0145951.
PR	21-MAY-1999;	99US-0135124.		PR	02-AUG-1999;	99US-0146386.
PR	24-MAY-1999;	99US-0135353.		PR	02-AUG-1999;	99US-0146388.
PR	25-MAY-1999;	99US-0135629.		PR	02-AUG-1999;	99US-0146389.
PR	27-MAY-1999;	99US-0136021.		PR	03-AUG-1999;	99US-0147038.
PR	28-MAY-1999;	99US-0136392.		PR	04-AUG-1999;	99US-0147204.
PR	01-JUN-1999;	99US-0136782.		PR	04-AUG-1999;	99US-0147302.
PR	03-JUN-1999;	99US-0137222.		PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.		PR	06-AUG-1999;	99US-0147260.
PR	07-JUN-1999;	99US-0137502.		PR	06-AUG-1999;	99US-0147303.
PR	08-JUN-1999;	99US-0137724.		PR	09-AUG-1999;	99US-0147416.
PR	10-JUN-1999;	99US-0138094.		PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.		PR	09-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	99US-0138847.		PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139119.		PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.		PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	99US-0139453.		PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139492.		PR	16-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.		PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.		PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.		PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.		PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.		PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.		PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139460.		PR	25-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139461.		PR	26-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139462.		PR	27-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139463.		PR	27-AUG-1999;	99US-0151065.
PR	21-JUN-1999;	99US-0139750.		PR	30-AUG-1999;	99US-0151066.
PR	22-JUN-1999;	99US-0139763.		PR	31-AUG-1999;	99US-0151080.
PR	23-JUN-1999;	99US-0139899.		PR	01-SEP-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.		PR	07-SEP-1999;	99US-0151438.
PR	24-JUN-1999;	99US-0140354.		PR	10-SEP-1999;	99US-0151930.
PR	28-JUN-1999;	99US-0140695.		PR	15-SEP-1999;	99US-0152363.
PR	29-JUN-1999;	99US-0140823.		PR	16-SEP-1999;	99US-0153070.
PR	30-JUN-1999;	99US-0140901.		PR	20-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	99US-0141287.		PR	22-SEP-1999;	99US-0154018.
PR	02-JUL-1999;	99US-0141842.		PR	23-SEP-1999;	99US-0154039.
PR	02-JUL-1999;	99US-0142154.		PR	24-SEP-1999;	99US-0154779.
PR	06-JUL-1999;	99US-0142055.		PR	28-SEP-1999;	99US-0155139.
PR	08-JUL-1999;	99US-0142390.		PR	29-SEP-1999;	99US-0155486.
PR	09-JUL-1999;	99US-0142803.		PR	04-OCT-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.		PR	05-OCT-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142277.		PR	07-OCT-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.		PR	08-OCT-1999;	99US-0157117.
PR	13-JUL-1999;	99US-0143624.		PR	12-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.		PR	13-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.		PR	13-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.		PR	14-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.		PR	14-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.		PR	14-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.		PR	14-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.		PR	18-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.		PR	21-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.		PR	21-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.		PR	21-OCT-1999;	99US-0159331.
PR	21-JUL-1999;	99US-0144632.		PR	21-OCT-1999;	99US-0159637.
PR	21-JUL-1999;	99US-0144884.		PR	21-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.		PR	21-OCT-1999;	99US-0159584.
PR	22-JUL-1999;	99US-0145086.		PR	21-OCT-1999;	99US-0160741.
PR	22-JUL-1999;	99US-0145088.		PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.		PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.		PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.		PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.		PR	21-OCT-1999;	99US-0160815.

```
PR   22-OCT-1999;  99US-0160980.
PR   22-OCT-1999;  99US-0160981.
PR   22-OCT-1999;  99US-0160989.
PR   25-OCT-1999;  99US-0161404.
PR   25-OCT-1999;  99US-0161405.
PR   25-OCT-1999;  99US-0161406.
PR   26-OCT-1999;  99US-0161359.
PR   26-OCT-1999;  99US-0161360.
PR   26-OCT-1999;  99US-0161361.
PR   28-OCT-1999;  99US-0161920.
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.
```

Query Match   50.9%;   Score 88.5; DB 21; Length 133;
Best Local Similarity   66.7%;   Pred. No. 0.001;
Matches 20; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

```
Qy    1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
          || |||| | || | |||| ||| | |: | |:
Db  105 gggYygg-gggYygggggYggYgggdggggf 133
```

RESULT 15
G30159

```
ID   G30159 standard; Protein; 158 AA.
XX
AC   G30159;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 36007.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   29-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   23-APR-1999;  99US-0130891.
PR   28-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
```

```
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
PR   27-MAY-1999;  99US-0136392.
PR   28-MAY-1999;  99US-0136782.
PR   01-JUN-1999;  99US-0137222.
PR   03-JUN-1999;  99US-0137528.
PR   04-JUN-1999;  99US-0137502.
PR   07-JUN-1999;  99US-0137724.
PR   08-JUN-1999;  99US-0138094.
PR   10-JUN-1999;  99US-0138840.
PR   14-JUN-1999;  99US-0138847.
PR   16-JUN-1999;  99US-0139119.
PR   16-JUN-1999;  99US-0139452.
PR   17-JUN-1999;  99US-0139453.
PR   18-JUN-1999;  99US-0139492.
PR   18-JUN-1999;  99US-0139454.
PR   18-JUN-1999;  99US-0139455.
PR   18-JUN-1999;  99US-0139456.
PR   18-JUN-1999;  99US-0139457.
PR   18-JUN-1999;  99US-0139458.
PR   18-JUN-1999;  99US-0139459.
PR   18-JUN-1999;  99US-0139460.
PR   18-JUN-1999;  99US-0139461.
PR   18-JUN-1999;  99US-0139462.
PR   18-JUN-1999;  99US-0139463.
PR   18-JUN-1999;  99US-0139750.
PR   18-JUN-1999;  99US-0139763.
PR   21-JUN-1999;  99US-0139817.
PR   22-JUN-1999;  99US-0139899.
PR   23-JUN-1999;  99US-0140353.
PR   23-JUN-1999;  99US-0140354.
PR   24-JUN-1999;  99US-0140695.
PR   28-JUN-1999;  99US-0140823.
PR   29-JUN-1999;  99US-0140991.
PR   30-JUN-1999;  99US-0141287.
PR   01-JUL-1999;  99US-0141842.
PR   01-JUL-1999;  99US-0142154.
PR   02-JUL-1999;  99US-0142055.
PR   06-JUL-1999;  99US-0142390.
PR   08-JUL-1999;  99US-0142803.
PR   09-JUL-1999;  99US-0142920.
PR   12-JUL-1999;  99US-0142977.
PR   13-JUL-1999;  99US-0143542.
PR   14-JUL-1999;  99US-0143624.
PR   15-JUL-1999;  99US-0144005.
PR   16-JUL-1999;  99US-0144085.
PR   16-JUL-1999;  99US-0144086.
PR   19-JUL-1999;  99US-0144325.
PR   19-JUL-1999;  99US-0144331.
PR   19-JUL-1999;  99US-0144332.
PR   19-JUL-1999;  99US-0144333.
PR   19-JUL-1999;  99US-0144334.
PR   19-JUL-1999;  99US-0144335.
PR   20-JUL-1999;  99US-0144352.
PR   20-JUL-1999;  99US-0144632.
PR   20-JUL-1999;  99US-0144884.
PR   21-JUL-1999;  99US-0144814.
PR   21-JUL-1999;  99US-0145086.
PR   21-JUL-1999;  99US-0145088.
PR   22-JUL-1999;  99US-0145085.
PR   22-JUL-1999;  99US-0145087.
PR   22-JUL-1999;  99US-0145089.
PR   22-JUL-1999;  99US-0145192.
PR   23-JUL-1999;  99US-0145145.
PR   23-JUL-1999;  99US-0145218.
PR   26-JUL-1999;  99US-0145224.
PR   27-JUL-1999;  99US-0145276.
PR   27-JUL-1999;  99US-0145913.
PR   27-JUL-1999;  99US-0145918.
PR   27-JUL-1999;  99US-0145919.
```

PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match   50.9%;   Score 88.5;   DB 21;   Length 158;
Best Local Similarity 66.7%;   Pred. No. 0.0012;
Matches 20; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
        || ||| ||| || ||| || ||| || ||:
Db 130 gggyygg-gggcyyggcggcyyggdgggg$ 158

Search completed: April 24, 2001, 16:38:22
Job time: 422 sec

PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 30-AUG-1999; 99US-0151066.
PR 31-AUG-1999; 99US-0151080.
PR 01-SEP-1999; 99US-0151303.
PR 07-SEP-1999; 99US-0151438.
PR 10-SEP-1999; 99US-0151930.
PR 13-SEP-1999; 99US-0152363.
PR 15-SEP-1999; 99US-0153070.
PR 16-SEP-1999; 99US-0153758.
PR 20-SEP-1999; 99US-0154018.
PR 22-SEP-1999; 99US-0154039.
PR 23-SEP-1999; 99US-0154779.
PR 24-SEP-1999; 99US-0155139.
PR 28-SEP-1999; 99US-0155486.
PR 29-SEP-1999; 99US-0155659.
PR 04-OCT-1999; 99US-0156458.
PR 05-OCT-1999; 99US-0156596.
PR 06-OCT-1999; 99US-0157117.
PR 07-OCT-1999; 99US-0157753.
PR 08-OCT-1999; 99US-0157865.
PR 12-OCT-1999; 99US-0158029.
PR 13-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 14-OCT-1999; 99US-0159293.
PR 14-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 18-OCT-1999; 99US-0159329.
PR 21-OCT-1999; 99US-0159330.
PR 21-OCT-1999; 99US-0159331.
PR 21-OCT-1999; 99US-0159637.
PR 21-OCT-1999; 99US-0159638.
PR 22-OCT-1999; 99US-0159584.
PR 22-OCT-1999; 99US-0160741.
PR 25-OCT-1999; 99US-0160757.
PR 25-OCT-1999; 99US-0160768.
PR 26-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    April 24, 2001, 16:36:25 ; Search time 62.39 Seconds
                                       (without alignments)
                                       9.237 Million cell updates/sec

Title:         US-09-340-736-8
Perfect score: 174
Sequence:      1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

rched:         185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters:    185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     Issued_Patents_AA:*
          1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
          2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
          3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
          4: /cgn2_5/ptodata/2/iaa/6B_COMB.pep:*
          5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
          6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                             SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	174	100.0	30	2	US-08-911-364-8	Sequence 8, Appli
2	109	62.6	123	2	US-07-609-716-103	Sequence 103, App
3	109	62.6	123	4	US-08-475-411A-103	Sequence 103, App
4	109	62.6	123	4	US-08-478-029A-103	Sequence 103, App
5	109	62.6	159	1	US-07-609-716-104	Sequence 104, App
6	109	62.6	159	4	US-08-475-411A-104	Sequence 104, App
7	109	62.6	159	4	US-08-478-029A-104	Sequence 104, App
8	96	55.2	111	4	US-08-963-168C-15	Sequence 15, Appl
9	96	55.2	116	4	US-08-963-168C-13	Sequence 13, App-
10	96	55.2	131	4	US-08-963-168C-8	Sequence 8, Appli
11	96	55.2	136	4	US-08-963-168C-6	Sequence 6, Appli
12	93	53.4	745	2	US-09-010-928B-28	Sequence 28, Appl
13	93	53.4	870	2	US-09-010-928B-2	Sequence 2, Appli
14	93	53.4	907	2	US-09-010-928B-4	Sequence 4, Appli
15	91	52.3	106	4	US-08-963-168C-14	Sequence 14, Appl
16	91	52.3	126	4	US-08-963-168C-7	Sequence 7, Appli
17	88	50.6	718	2	US-08-425-069-2	Sequence 2, Appli
18	88	50.6	718	2	US-08-317-844B-2	Sequence 2, Appli
19	88	50.6	747	4	US-09-034-177-3	Sequence 3, Appli
20	87.5	50.3	112	4	US-08-963-168C-16	Sequence 16, Appli
21	87.5	50.3	132	4	US-08-963-168C-9	Sequence 9, Appli
22	85	48.9	235	2	US-08-529-190B-1	Sequence 1, Appli
23	82	47.1	341	2	US-08-538-711A-8	Sequence 8, Appli
24	82	47.1	353	2	US-08-538-711A-7	Sequence 7, Appli
25	81	46.6	40	3	US-08-815-190A-17	Sequence 17, Appl
26	79	45.4	1345	2	US-08-977-767-3	Sequence 3, Appli
27	77.5	44.5	832	1	US-08-209-747-2	Sequence 2, Appli

28	77.5	44.5	832	1	US-08-458-298-2	Sequence 2, Appli
29	76.5	44.0	731	2	US-08-911-364-1	Sequence 1, Appli
30	76.5	44.0	792	2	US-08-678-039A-40	Sequence 40, Appl
31	76	43.7	124	4	US-08-789-333F-41	Sequence 41, Appl
32	76	43.7	127	4	US-08-789-333F-43	Sequence 43, Appl
33	76	43.7	173	4	US-08-789-333F-42	Sequence 42, Appl
34	76	43.7	177	4	US-08-789-333F-44	Sequence 44, Appl
35	75	43.1	546	2	US-09-067-351-1	Sequence 1, Appli
36	74.5	42.8	738	3	US-08-664-038A-3	Sequence 3, Appli
37	74	42.5	660	3	US-08-058-489-18	Sequence 18, Appl
38	74	42.5	1958	1	US-09-058-489-91	Sequence 91, Appl
39	71	40.8	50	1	US-07-945-283-2	Sequence 2, Appli
40	70.5	40.5	50	1	US-08-209-747-27	Sequence 27, Appl
41	70.5	40.5	50	1	US-08-458-298-27	Sequence 27, Appl
42	70.5	40.5	51	1	US-08-209-747-20	Sequence 20, Appl
43	70.5	40.5	51	1	US-08-458-298-20	Sequence 20, Appl
44	70.5	40.5	385	1	US-08-891-254-3	Sequence 3, Appli
45	70.5	40.5	385	2	US-08-819-539-3	Sequence 3, Appli

```
                             ALIGNMENTS

RESULT  1
US-08-911-364-8
 ; Sequence 8, Application US/08911364
 ; Patent No.5969106
 ; GENERAL INFORMATION:
 ;  APPLICANT: ROTHSTEIN, Aser
 ;  APPLICANT: KEELY, Fred W.
 ;  APPLICANT: ROTHSTEIN, Steven J.
 ;  TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
 ;  TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
 ;  NUMBER OF SEQUENCES: 8
 ;  CORRESPONDENCE ADDRESS:
 ;   ADDRESSEE: FOLEY & LARDNER
 ;   STREET: 3000 K Street, N.W.
 ;   CITY: Washington
 ;   STATE: D.C.
 ;   COUNTRY: U.S.A.
 ;   ZIP: 20007-5109
 ;  COMPUTER READABLE FORM:
 ;   MEDIUM TYPE: Floppy disk
 ;   COMPUTER: IBM PC Compatible
 ;   OPERATING SYSTEM: PC-DOS/MS-DOS
 ;   SOFTWARE: PatentIn Release #1.0, Version #1.30
 ;  CURRENT APPLICATION DATA:
 ;   APPLICATION NUMBER: US/08/911,364
 ;   FILING DATE: 07-AUG-1997
 ;   CLASSIFICATION: 530
 ;  PRIOR APPLICATION DATA:
 ;   APPLICATION NUMBER: US 60/023,552
 ;   FILING DATE: 07-AUG-1996
 ;  ATTORNEY/AGENT INFORMATION:
 ;   NAME: Bent, Stephen A.
 ;   REGISTRATION NUMBER: 29,768
 ;   REFERENCE/DOCKET NUMBER: 041082/0104
 ;  TELECOMMUNICATION INFORMATION:
 ;   TELEPHONE: (202) 672-5300
 ;   TELEFAX: (202) 672-5399
 ;  INFORMATION FOR SEQ ID NO: 8:
 ;   SEQUENCE CHARACTERISTICS:
 ;    LENGTH: 30 amino acids
 ;    TYPE: amino acid
 ;    STRANDEDNESS:
 ;    TOPOLOGY: linear
 ;   MOLECULE TYPE: peptide
 US-08-911-364-8

Query Match         100.0%;  Score 174;  DB 2;  Length 30;
Best Local Similarity  100.0%;  Pred. No. 5.2e-15;
Matches   30;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy      1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
          |||||||||||||||||||||||||||||
Db      1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
```

RESULT 2
US-07-609-716-103
; Sequence 103, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
;   APPLICANT: Ferrari, Franco A.
;   APPLICANT: Cappello, Joseph
;   TITLE OF INVENTION: Functional Recombinantly Prepared
;   TITLE OF INVENTION: Synthetic Protein Polymer
;   NUMBER OF SEQUENCES: 118
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: Four Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: US
;   ZIP: 94111
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/609,716
;   FILING DATE: 06-NOV-1990
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Rowland, Bertram I
;   REGISTRATION NUMBER: 20015
;   REFERENCE/DOCKET NUMBER: A-55186-3/BIR
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-781-1989
;   TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 103:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 123 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-07-609-716-103

Query Match   62.6%; Score 109; DB 1; Length 123;
Best Local Similarity 62.1%; Pred. No. 8.5e-07; Indels 0;
Matches 18; Conservative 6; Mismatches 5; Gaps 0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
          ||:| ||:|| ||:|| ||:|| ||:||:|
Db   35 GGVGVGGVGVGGVGVGGVGVGGVGVGGVGVGGVG 63
```

RESULT 3
US-08-475-411A-103
; Sequence 103, Application US/08475411A
; Patent No. 6140072
; GENERAL INFORMATION:
;    APPLICANT: Ferrari, Franco A.
;    APPLICANT: Cappello, Joseph
;    TITLE OF INVENTION: Functional Recombinantly Prepared
;    TITLE OF INVENTION: Synthetic Protein Polymer
;    NUMBER OF SEQUENCES: 119
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;       STREET: Four Embarcadero Center, Suite 3400
;       CITY: San Francisco
;       STATE: CA

---

;       COUNTRY: US
;       ZIP: 94111
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/475,411A
;       FILING DATE: 07-JUN-1995
;       CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 07/609,716
;       FILING DATE: 06-NOV-1990
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 07/269,429
;       FILING DATE: 09-NOV-1988
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 07/114,618
;       FILING DATE: 29-OCT-1987
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 06/927,258
;       FILING DATE: 04-NOV-1986
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Trecartin, Richard F.
;       REGISTRATION NUMBER: 31,801
;       REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 415-781-1989
;       TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 103:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH: 123 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;       MOLECULE TYPE: peptide
US-08-475-411A-103

Query Match        52.6%; Score 109; DB 4; Length 123;
Best Local Similarity  52.1%; Pred. No. 8.5e-07; Indels  0;
Matches 18; Conservative  6; Mismatches  5; Gaps  0;

Qy      1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
          ||:|  ||:||  ||:||  ||:||  ||:||:|
Db     35 GGVGVGGVGVGGVGVGGVGVGGVGVGGVGVGGVG 63
```

RESULT 4
US-08-478-029A-103
; Sequence 103, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
;   APPLICANT: Ferrari, Franco A.
;   APPLICANT: Cappello, Joseph
;   TITLE OF INVENTION: Functional Recombinantly Prepared
;   TITLE OF INVENTION: Synthetic Protein Polymer
;   NUMBER OF SEQUENCES: 119
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: Four Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: US
;   ZIP: 94111
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/478,029A
```

; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
US-08-478-029A-103

Query Match      62.6%;  Score 109;  DB 4;  Length 123;
Best Local Similarity   62.1%;  Pred. No. 8.5e-07;
Matches  18;  Conservative  6;  Mismatches 5;  Indels 0;  Gaps 0;

Qy  1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
       ||:||:||:||:||:||:||:||:|||
Db 35 GGVGVGGVGVGGVGVGGVGVGGVGVGGVG 63

RESULT 5
US-07-509-716-104
; Sequence 104, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249

INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-609-716-104

Query Match      62.6%;  Score 109;  DB 1;  Length 159;
Best Local Similarity   62.1%;  Pred. No. 1.1e-06;
Matches  18;  Conservative  6;  Mismatches 5;  Indels 0;  Gaps 0;

Qy  1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
       ||:||:||:||:||:||:||:||:|||
Db 53 GGVGVGGVGVGGVGVGGVGVGGVGVGGVG 81

RESULT 6
US-08-475-411A-104
; Sequence 104, Application US/08475411A
; Patent No. 6140072
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-475-411A-104

```
Query Match        62.6%;   Score 109;   DB 4;   Length 159;
Best Local Similarity  52.1%;  Pred. No. 1.1e-06;  Indels  0;  Gaps  0;
Matches  18;  Conservative  6;  Mismatches  5;

Qy   1  GGLGYGGLGYGGLGYGGLGYGGLGYGGLG  29
         ||:| ||:| ||:| ||:| ||:| ||:|
Db  53  GGVGVGVGVGVGVGVGVGVGVGVGVGVG  81
```

RESULT 7
US-08-478-029A-104
; Sequence 104, Application US/08478029A
; Patent No. 618348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-478-029A-104

```
Query Match        62.6%;   Score 109;   DB 4;   Length 159;
Best Local Similarity  52.1%;  Pred. No. 1.1e-06;  Indels  0;  Gaps  0;
Matches  18;  Conservative  6;  Mismatches  5;

Qy   1  GGLGYGGLGYGGLGYGGLGYGGLGYGGLG  29
         ||:| ||:| ||:| ||:| ||:| ||:|
Db  53  GGVGVGVGVGVGVGVGVGVGVGVGVGVG  81
```

RESULT 8
US-08-963-168C-15
; Sequence 15, Application US/08963168C
; Patent No. 6127166
; GENERAL INFORMATION:
; APPLICANT: Bayley, Hagan
; APPLICANT: Cao, Quiping
; APPLICANT: Wang, Yunjaun
; TITLE OF INVENTION: MOLLUSCAN LIGAMENT POLYPEPTIDES
; TITLE OF INVENTION: AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,168C
; FILING DATE: 03-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/059001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 111 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-963-168C-15

```
Query Match        55.2%;   Score 96;   DB 4;   Length 111;
Best Local Similarity  50.0%;  Pred. No. 2.6e-05;  Indels  7;  Gaps  1;
Matches  17;  Conservative  6;  Mismatches  7;

Qy   1  GGLGYGGLGYGGLGYGGLGYGGL---GYGGLGY  30
         ||:| ||:||:| |:|| |:|| |||
Db  78  GGYGSGNAGFGGMGGGNAGFGGMGGGGFGSKGY  111
```

RESULT 9
US-08-963-168C-13
; Sequence 13, Application US/08963168C
; Patent No. 6127166
; GENERAL INFORMATION:
; APPLICANT: Bayley, Hagan
; APPLICANT: Cao, Quiping
; APPLICANT: Wang, Yunjaun
; TITLE OF INVENTION: MOLLUSCAN LIGAMENT POLYPEPTIDES
; TITLE OF INVENTION: AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/963,168C
FILING DATE: 03-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, Peter J.
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07917/059001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-963-168C-13

Query Match 55.2%; Score 96; DB 4; Length 116;
Best Local Similarity 50.0%; Pred. No. 2.7e-05; Indels 4; Gaps 1;
Matches 17; Conservative 6; Mismatches 7; Indels 4; Gaps 1;

Qy 1 GGLGYGGLGYGGLGYGGLGYGGL---GYGGLGY 30
       ||:|| |:||:| | |:||:| |:|| ||
Db 83 GGMGGGNAGFGGMGGGNAGFGGMGGGQGGFGGKGY 116

RESULT 10:
US-08-963-168C-8
; Sequence 8, Application US/08963168C
; Patent No. 6127166
; GENERAL INFORMATION:
; APPLICANT: Bayley, Hagan
; APPLICANT: Cao, Quiping
; APPLICANT: Wang, Yunjaun
; TITLE OF INVENTION: MOLLUSCAN LIGAMENT POLYPEPTIDES
; TITLE OF INVENTION: AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,168C
; FILING DATE: 03-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/059001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal

US-08-963-168C-8

Query Match 55.2%; Score 96; DB 4; Length 131;
Best Local Similarity 50.0%; Pred. No. 3e-05; Indels 4; Gaps 1;
Matches 17; Conservative 6; Mismatches 7; Indels 4; Gaps 1;

Qy 1 GGLGYGGLGYGGLGYGGLGYGGL---GYGGLGY 30
       ||:|| |:||:| | |:||:| |:|| ||
Db 98 GGMGGGNAGFGGMGGGNAGFGGMGGGQGGFGGKGY 131

RESULT 11
US-08-963-168C-6
; Sequence 6, Application US/08963168C
; Patent No. 6127166
; GENERAL INFORMATION:
; APPLICANT: Bayley, Hagan
; APPLICANT: Cao, Quiping
; APPLICANT: Wang, Yunjaun
; TITLE OF INVENTION: MOLLUSCAN LIGAMENT POLYPEPTIDES
; TITLE OF INVENTION: AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,168C
; FILING DATE: 03-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/059001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 136 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; FEATURE:
; OTHER INFORMATION: This translation is for SEQ ID NOs:1 & 2.
US-08-963-168C-6

Query Match 55.2%; Score 96; DB 4; Length 136;
Best Local Similarity 50.0%; Pred. No. 3.1e-05; Indels 4; Gaps 1;
Matches 17; Conservative 6; Mismatches 7; Indels 4; Gaps 1;

Qy 1 GGLGYGGLGYGGLGYGGLGYGGL---GYGGLGY 30
       ||:|| |:||:| | |:||:| |:|| ||
Db 103 GGMGGGNAGFGGMGGGNAGFGGMGGGQGGFGGKGY 136

RESULT 12
US-09-010-928B-28
; Sequence 28, Application US/09010928B
; Patent No. 594099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V

; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 745 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-010-928B-28

Query Match   53.4%; Score 93; DB 2; Length 745;
Best Local Similarity   62.1%; Pred. No. 0.00035;
Matches 18; Conservative   1; Mismatches 10; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
        || || |||| || || || || || ||||| |
Db 702 GGAGPGGAGPGGVGLGGAGRGGAGRGGAG 730

RESULT 13
US-09-010-928B-2
; Sequence 2, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 870 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-010-928B-2

Query Match   53.4%; Score 93; DB 2; Length 870;
Best Local Similarity   62.1%; Pred. No. 0.00041;
Matches 18; Conservative   1; Mismatches 10; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
        || || |||| || || || || || ||||| |
Db 828 GGAGPGGAGPGGVGLGGAGRGGAGRGGAG 855

RESULT 14
US-09-010-928B-4
; Sequence 4, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-010-928B-4

Query Match   53.4%; Score 93; DB 2; Length 907;
Best Local Similarity   62.1%; Pred. No. 0.00042;
Matches 18; Conservative   3; Mismatches   8; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
        || || |||| || || || || || ||| |
Db 728 GGVGSGSGSGPGGVPGGYGPGPGGSGSGGVG 756

RESULT 15
US-08-963-168C-14
; Sequence 14, Application US/08963168C
; Patent No. 6127166
; GENERAL INFORMATION:
; APPLICANT: Bayley, Hagan
; APPLICANT: Cao, Quiping
; APPLICANT: Wang, Yunjaun

TITLE OF INVENTION: MOLLUSCAN LIGAMENT POLYPEPTIDES
TITLE OF INVENTION: AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
    ADDRESSEE: Fish & Richardson P.C.
    STREET: 225 Franklin Street
    CITY: Boston
    STATE: MA
    COUNTRY: US
    ZIP: 02110-2804
COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette
    COMPUTER: IBM Compatible
    OPERATING SYSTEM: Windows95
    SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/963,168C
    FILING DATE: 03-NOV-1997
ATTORNEY/AGENT INFORMATION:
    NAME: Fasse, Peter J.
    REGISTRATION NUMBER: 32,983
    REFERENCE/DOCKET NUMBER: 07917/059001
TELECOMMUNICATION INFORMATION:
    TELEPHONE: 617/542-5070
    TELEFAX: 617/542-8906
    TELEX: 200154
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
    LENGTH: 106 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
    MOLECULE TYPE: protein
US-08-963-168C-14

Query Match   52.3%; Score 91; DB 4; Length 106;
Best Local Similarity   50.0%; Pred. No. 9.6e-05;
Matches 17; Conservative   6; Mismatches   7; Indels   4; Gaps   1;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGL---GYGGLGY 30
            ||:| |:||:|| |:||| |:||| |
Db   73 GGMGSGMGGFGGMGGGNAGFGGMGGQGGFGGKGY 106

rch completed: April 24, 2001, 16:36:25
    time: 305 sec

OM protein - protein search, using sw model

```
                 GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.
```

Run on: April 24, 2001, 16:42:04 ; Search time 74.56 seconds
                                    (without alignments)
                                   27.651 million cell updates/sec

Title: US-09-340-736-8
Perfect score: 174
Sequence: 1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30

Scoring table: BLOSUM62
                GapOp 10.0 , Gapext 0.5

Searched: 198801 seqs. 68722935 residues   198801

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR_67:*
             1: Pir1:*
             2: Pir2:*
             3: Pir3:*
             4: Pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	DB ID	Description
1	160	92.0	119	1	B45051	lamprin 1 precursor
2	146	83.9	139	1	C45051	lamprin 2 precursor
3	140	80.5	151	2	S04515	chorion class B pr
4	135	78.2	168	2	C24255	chorion class B pr
5	122.5	70.4	110	2	S01420	chorion class B pr
6	122	70.1	860	1	EAMS	elastin precursor
7	122	70.1	864	1	EART	elastin precursor
8	119	68.4	171	2	JBA041	chorion class B pr
9	115.5	66.4	119	2	JAA092	chorion class A pr
10	109.5	62.9	155	1	KRCHS	keratin, scale - c
11	106.5	61.2	259	2	T15126	hypothetical prote
12	105	60.1	91	2	S01421	chorion class B pr
13	104.5	60.1	119	2	S04513	chorion class B pr
14	104	59.5	89	2	T25923	hypothetical prote
15	103.5	59.5	513	2	S08381	keratin, 58K type
16	101.5	57.5	80	2	B45466	glycine/tyrosine-r
17	100	57.5	643	1	KRHU2	keratin 1, type II
18	99	56.9	783	2	T35389	keratin, type II
19	98	56.3	73	2	A60136	probable serine-th
20	97	55.7	94	2	T22579	keratin, scale (cl
21	97	55.7	168	2	S12312	hypothetical prote
22	97	55.5	126	2	S04514	glycine-rich RNA-b
23	96.5	55.5	784	2	A26601	chorion class B pr
24	95	54.6	199	2	A19324	elastin precursor
25	95	54.3	102	2	T20557	chorion class C pr
26	94	54.3	128	2	J01002	hypothetical prote
27	94	54.0	165	2	S41773	keratin, claw - ch
28	94	54.0	265	2	S59529	glycine-rich glyci
29	93	53.4	129	2	A24255	chorion class A pr
30	93	53.4		2	KRXL	keratin 3, type I,
31	92	52.9		2	S40777	heterogeneous ribo
32	92	52.9		2	T08080	probable myrosinas
33	91.5	52.6		2	T02026	glycine-rich prote
34	91.5	52.6		1	B75374	conserved hypothet
35	91.5	52.6		1	S40778	ribonucleoprotein
36	91	52.3		1	KRMS2	keratin, type II c
37	90.5	52.0		1	S23060	early chorion prot
38	90.5	52.0		1	T19803	hypothetical prote
39	90.5	52.0		1	T19804	hypothetical prote
40	90	51.7		2	A25297	chorion class B pr
41	90	51.7		2	T03583	glycine-rich RNA-b
42	90	51.7		2	S42070	early chorion prot
43	90	51.7		2	S38331	glycine-rich RNA-b
44	90	51.7		2	S42071	early chorion prot
45	90	51.7		2	S23463	cuticular protein

ALIGNMENTS

RESULT   1
B45051
lamprin 1 precursor - sea lamprey
N;Alternate names: lamprin L-1.8-10
C;Species: Petromyzon marinus (sea lamprey)
C;Date: 30-Apr-1993 #sequence_revision 30-Apr-1993 #text_change 21-Jul-2000
C;Accession: B45051
R;Robson, P.; Wright, G.M.; Sitarz, E.; Maiti, A.; Rawat, M.; Youson, J.H.; Keeley, F
J. Biol. Chem. 268, 1440-1447, 1993
A;Title: Characterization of lamprin, an unusual matrix protein from lamprey cartilag
A;Reference number: A45051; MUID:93123269
A;Accession: B45051
A;Molecule type: mRNA
A;Residues: 1-119 <ROB>
A;Cross-references: GB:L05926; NID:g213211; PIDN:AAA49270.1; PID:g213212
A;Note: sequence extracted from NCBI backbone (NCBIP:122168)
C;Comment: This sequence is encoded by a different gene than C45051; we have arbitrar
C;Superfamily: lamprin
C;Keywords: cartilage; extracellular matrix
P;1-19/Domain: signal sequence #status predicted <SIG>
F;20-119/Product: lamprin 1 #status predicted <MAT>

```
Query Match             92.0%;  Score 160; DB 1; Length 119;
Best Local Similarity   93.3%;  Pred. No. 1.8e-11;
Matches    28;  Conservative   0; Mismatches   2; Indels    0; Gaps    0;

Qy     1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
           |||||||||||||||||||| |||||||
Db    41 GGLGYGGLGYGGLGYGGLGYGGLGVAGLGYGGLGY 70
```

RESULT   2
C45051
lamprin 2 precursor, long splice form - sea lamprey
N;Alternate names: lamprin L-0.9-12
N;Contains: lamprin 2 precursor, short splice form (lamprin L-0.9-10)
C;Species: Petromyzon marinus (sea lamprey)
C;Date: 30-Apr-1993 #sequence_revision 30-Apr-1993 #text_change 22-Jun-1999
C;Accession: C45051; B45051
R;Robson, P.; Wright, G.M.; Sitarz, E.; Maiti, A.; Rawat, M.; Youson, J.H.; Keeley, F
J. Biol. Chem. 268, 1440-1447, 1993
A;Title: Characterization of lamprin, an unusual matrix protein from lamprey cartilag
A;Reference number: A45051; MUID:93123269
A;Accession: C45051
A;Molecule type: mRNA
A;Residues: 1-139 <ROB>
A;Cross-references: GB:L05925; NID:g213209; PIDN:AAA49269.1; PID:g213210
A;Note: sequence extracted from NCBI backbone (NCBIP:122170)
A;Accession: A45051
A;Molecule type: mRNA
A;Residues: 1-85,105-139 <RO2>

A;Cross-references: GB:L05924; NID:g213207; PIDN:AAA49268.1; PID:g213208
A;Note: sequence extracted from NCBI backbone (NCBIP:1122166)
C;Comment: These forms are encoded by a different gene than B45051; we have arbitrarily
C;Superfamily: lamprin
C;Keywords: alternative splicing; cartilage; extracellular matrix
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-139/Product: lamprin 2, long splice form #status predicted <MAT1>
F;20-85/Product: lamprin 2, short splice form #status predicted <MAT2>

Query Match   83.9%; Score 146; DB 1; Length 139;
Best Local Similarity   86.7%; Pred. No. 7e-10;
Matches   26; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
          |||||||||||||||| || |||||||||
Db   42 GGLGYGGLGYGGLGYVAGLGVAGLGYGGLGY 71

RESULT 3
S04515
chorion class B protein precursor (clone 410) - silkworm (fragment)
C;Species: Bombyx mori (silkworm)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 15-Nov-1996
C;Accession: S04515
R;Tsitilou, S.G.; Rodakis, G.C.; Alexopoulou, M.; Kafatos, F.C.; Ito, K.; Iatrou, K.
EMBO J. 2, 1845-1852, 1983
A;Title: Structural features of B family chorion sequences in the silkmoth Bombyx mori,
A;Reference number: S04515
A;Accession: S04515
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-151 <TSI>
C;Superfamily: chorion class A protein pc292
F;1-4/Domain: signal sequence (fragment) #status predicted <SIG>
F;5-151/Product: chorion class B protein #status predicted <MAT>

Query Match   80.5%; Score 140; DB 2; Length 151;
Best Local Similarity   89.3%; Pred. No. 3.4e-09;
Matches   25; Conservative   0; Mismatches   3; Indels   0; Gaps   0;

Qy   2 GLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
          || ||||||||||||||||||
Db   8 GRGCGGRGYGGLGYGGLGYGGLGYGGLGYGGLG 35

RESULT 4
C24255
chorion class B protein L11 precursor - silkworm
C;Species: Bombyx mori (silkworm)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 01-Dec-2000
A;Accession: C24255; D24255; S05914
R;Spoerel, N.; Nguyen, H.T.; Kafatos, F.C.
J. Mol. Biol. 190, 23-35, 1986
A;Title: Gene regulation and evolution in the chorion locus of Bombyx mori. Structural a
A;Reference number: A92929; MUID:87060979
A;Accession: C24255
A;Molecule type: DNA
A;Residues: 1-168 <SPO>
A;Cross-references: GB:X15557; GB:X15557; GB:X04028; GB:X04029; GB:X04030; GB:X04031; GB:X04032; GB
A;Accession: D24255
A;Residues: 1-18,'V',20,'S',22-27,35,'R',37-50,'L',52-114,'D',116-130,'F',132,'Y',134-13
R;Spoerel, N.A.; Nguyen, H.T.; Eickbush, T.H.; Kafatos, F.C.
J. Mol. Biol. 209, 1-19, 1989
A;Title: Gene evolution and regulation in the chorion complex of Bombyx mori. Hybridizat
A;Accession: S05914
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-17 <SP3>

A;Cross-references: EMBL:X15558; NID:g5771; PID:g5774
A;Note: this sequence was submitted to the EMBL Data Library, Jun-1989
C;Genetics:
A;Map position: 2
C;Superfamily: chorion class A protein pc292
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-168/Product: chorion class B protein L11 #status predicted <MAT>

Query Match   78.2%; Score 136; DB 2; Length 168;
Best Local Similarity   85.7%; Pred. No. 1e-08;
Matches   24; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy   2 GLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
          | ||| |||||||||||||||| |
Db   25 GCGCGGRGYGGLGYGGLGYGGLGYGGYG 52

RESULT 5
S01420
chorion class B protein (clone M2410) - silkworm (fragment)
C;Species: Bombyx mori (silkworm)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 20-Aug-1999
C;Accession: S01420
R;Tsitilou, S.G.; Rodakis, G.C.; Alexopoulou, M.; Kafatos, F.C.; Ito, K.; Iatrou, K.
EMBO J. 2, 1845-1852, 1983
A;Title: Structural features of B family chorion sequences in the silkmoth Bombyx mor
A;Reference number: S01420; MUID:84057707
A;Accession: S01420
A;Molecule type: mRNA
A;Residues: 1-110 <TSI>
A;Cross-references: EMBL:X12837; NID:g5811; PIDN:CAA31322.1; PID:g5812
C;Superfamily: chorion class A protein pc292

Query Match   70.4%; Score 122.5; DB 2; Length 110;
Best Local Similarity   92.0%; Pred. No. 2e-07;
Matches   23; Conservative   0; Mismatches   1; Indels   1; Gaps   1;

Qy   5 YGGLGYGGLGYGGLGYGGLGYGGLG 29
          |||||||||||||||||| |||
Db   1 YGGLGYGGLGYGGLGYGGLG-GGCG 24

RESULT 6
EAMS
elastin precursor - mouse
N;Alternate names: tropoelastin
C;Species: Mus musculus (house mouse)
C;Date: 18-Aug-1995 #sequence_revision 16-Aug-1996 #text_change 22-Jun-1999
C;Accession: A55721
R;Wydner, K.S.; Sechler, J.L.; Boyd, C.D.; Passmore, H.C.
Genomics 23, 125-131, 1994
A;Title: Use of an intron length polymorphism to localize the tropoelastin gene to mo
A;Reference number: A55721; MUID:95130069
A;Accession: A55721
A;Molecule type: mRNA
A;Residues: 1-860 <WYD>
A;Cross-references: GB:U08210; NID:g473273; PIDN:AAA80155.1; PID:g473274
C;Genetics:
A;Map position: 5
C;Superfamily: elastin
C;Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-860/Product: elastin #status predicted <MAT>
F;850-855/Disulfide bonds: #status predicted

Query Match   70.1%; Score 122; DB 1; Length 860;
Best Local Similarity   82.8%; Pred. No. 1.7e-06;
Matches   24; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29

```
Db    733 GGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLG 761
          |||| |||| |||| |||| |||| |||| |||| ||||
Db    738 GGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLG 766
```

RESULT 7
EART
elastin precursor - rat
N:Alternate names: tropoelastin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jan-1991 #sequence_revision 16-Aug-1996 #text_change 22-Jun-1999
C:Accession: A36106; A30878; A36523; S02173; I54172; I68505
R:Pierce, R.A.; Deak, S.B.; Stolle, C.A.; Boyd, C.D.
Biochemistry 29, 9677-9683, 1990
A:Title: Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.
A:Reference number: A36106; MUID:91104868
A:Accession: A36106
A:Molecule type: mRNA
A:Residues: 1-864 <PIP>
A:Cross-references: GB:M60547; GB:J05292; NID:g207444; PIDN:AAA42269.1; PID:g207445
R:Deak, S.B.; Pierce, R.A.; Belsky, S.A.; Riley, D.J.; Boyd, C.D.
J. Biol. Chem. 263, 13504-13507, 1988
A:Title: Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.
A:Reference number: A30878; MUID:88330868
A:Accession: A30878
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 781-864 <DEA>
A:Cross-references: GB:V00035; NID:g207442; PIDN:AAA42268.1; PID:g207443
R:Franzblau, C.; Pratt, C.A.; Faris, B.; Colannino, N.M.; Offner, G.D.; Mogayzel Jr., P.
J. Biol. Chem. 264, 15115-15119, 1989
A:Title: Role of tropoelastin fragmentation in elastogenesis in rat smooth muscle cells
A:Reference number: A36523; MUID:89359927
A:Accession: A36523
A:Molecule type: protein
A:Residues: 22-31 <FRA>
R:Rich, C.B.; Foster, J.A.
Arch. Biochem. Biophys. 268, 551-556, 1989
A:Title: Characterization of rat heart tropoelastin.
A:Reference number: S02173; MUID:89117149
A:Accession: S02173
A:Molecule type: mRNA
A:Status: not compared with conceptual translation
A:Residues: ?1P, 369-545,548-764,770-864 <RIC>
A:Experimental source: heart
R:Pierce, R.A.; Alatawi, A.; Deak, S.B.; Boyd, C.D.
Genomics 12, 651-658, 1992
A:Title: Elements of the rat tropoelastin gene associated with alternative splicing.
A:Reference number: I54172; MUID:92241859
A:Accession: I54172
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 264-533 <RES>
A:Cross-references: GB:M63372; NID:g207455; PIDN:AAA42271.1; PID:g554527
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 558-864 <RE2>
A:Cross-references: GB:M63376; NID:g207459; PIDN:AAA42272.1; PID:g207462
C:Genetics:
A:Introns: 277/1; 292/1; 308/1; 339/1; 359/1; 419/1; 437/1; 467/1; 484/1; 601/1; 621/1;
A:Note: the list of introns may be incomplete
C:Superfamily: elastin
C:Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxylysine
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-864/Product: elastin #status predicted <MAT>
F:854-859/Disulfide bonds: #status predicted

Query Match   70.1%; Score 122; DB 1; Length 864;
Best Local Similarity   82.8%; Pred. No. 1.7e-06;
Matches   24; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29

```
Db    738 GGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLG 756
          |||| |||| |||| |||| |||| |||| |||| ||||
Db    738 GGLGAGGLGAGGLGAGGLGAGGLGAGGLGAGGLG 756
```

RESULT 8
JBAO4I
chorion class B protein pc401 precursor - polyphemus moth
C:Species: Antheraea polyphemus (polyphemus moth)
C:Date: 28-Feb-1980 #sequence_revision 28-Feb-1980 #text_change 15-Oct-1999
C:Accession: A03337
R:Jones, C.W.; Rosenthal, N.; Rodakis, G.C.; Kafatos, F.C.
Cell 18, 1317-1332, 1979
A:Title: Evolution of two major chorion multigene families as inferred from cloned cD
A:Reference number: A90786; MUID:80090072
A:Accession: A03337
A:Molecule type: mRNA
A:Residues: 1-171 <JON>
A:Cross-references: GB:V00077; GB:J01162; NID:g5630; PIDN:CAA23419.1; PID:g5631
C:Comment: This protein is one of many from the eggshell of the polyphemus moth.
C:Superfamily: chorion class A protein pc292
C:Keywords: egg shell; structural protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-171/Product: chorion class B protein pc401 #status predicted <MAT>

Query Match   68.4%; Score 119; DB 1; Length 171,
Best Local Similarity   75.9%; Pred. No. 7.5e-07;
Matches   22; Conservative   1; Mismatches   6; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
          ||||| ||||| |:| ||||| || |||| | ||
Db   136 GGLGYGGLGYEGVGYGYGLGYGGYGLGYGGCG 164

RESULT 9
JBAO92
chorion class A protein pc292 precursor - polyphemus moth (fragment)
C:Species: Antheraea polyphemus (polyphemus moth)
C:Date: 31-Aug-1980 #sequence_revision 31-Aug-1980 #text_change 15-Oct-1999
C:Accession: A03336
R:Tsitilou, S.G.; Regier, J.C.; Kafatos, F.C.
Nucleic Acids Res. 8, 1987-1997, 1980
A:Title: Selection and sequence analysis of a cDNA clone encoding a known chorion pro
A:Reference number: A03336; MUID:81053859
A:Accession: A03336
A:Molecule type: mRNA
A:Residues: 1-119 <TSI>
A:Cross-references: GB:V00078; NID:g5632; PIDN:CAA23420.1; PID:g5633
A:Note: partial direct sequencing of the protein revealed one difference from the abo
C:Superfamily: chorion class A protein pc292
C:Keywords: egg shell; structural protein
F:1-6/Domain: signal sequence (fragment) #status predicted <SIG>
F:7-119/Product: chorion class A protein pc292 #status predicted <MAT>

Query Match   66.4%; Score 115.5; DB 1; Length 119;
Best Local Similarity   68.6%; Pred. No. 1.3e-06;
Matches   24; Conservative   0; Mismatches   0; Indels   5; Gaps   1;

Qy   1 GGLGYGGL-----GYGGLGYGGLGYGGLGYGGLGY 30
          ||||||| |||| ||||| |||| |||| |
Db   10 GGLGLKGLAAPACGCGGLGYEGLGYGALGYDGLGY 44

RESULT 10
KRCHS
keratin, scale - chicken
N:Alternate names: beta keratin
C:Species: Gallus gallus (chicken)
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 22-Jun-1999
C:Accession: A02851; B60136; I50168
R:Gregg, K.; Wilton, S.D.; Parry, D.A.D.; Rogers, G.E.
EMBO J. 3, 175-178, 1984
```

A;Title: A comparison of genomic coding sequences for feather and scale keratins: struct
A;Reference number: A59988; MUID:84158528
A;Accession: A02851
A;Molecule type: DNA
A;Residues: 1-155 <GRE>
A;Cross-references: GB:X00315; NID:g63548; PIDN:CAA25084.1; PID:g63549
R;Wilton, S.D.; Crocker, L.A.; Rogers, G.E.
Biochim. Biophys. Acta 824, 201-208, 1985
A;Title: Isolation and characterisation of keratin mRNA from the scale epidermis of the
A;Reference number: A60136; MUID:85122780
A;Accession: B60136
A;Molecule type: mRNA
A;Residues: 90-155 <WIL>
A;Cross-references: GB:M25642; NID:g212231; PIDN:AAA48931.1; PID:g212232
A;Note: this mRNA was designated clone CSK12. See entry A60136 for clone CSK9
R;Sawyer, R.H.; Shanes R.S.;
Curr. Top. Dev. Biol 22, 235-253, 1987
A;Title: Expression of beta keratin genes during development of avian skin appendages.
A;Reference number: I50168; MUID:88003561
A;Accession: I50168
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
residues: 137-155 <SAW>
Cross-references: GB:M28422; NID:g211271; PIDN:AAA49631.1; PID:g211272
C;Superfamily: feather keratin
C;Keywords: duplication; epidermis; fibrous protein; horn; integument; tandem repeat
F;70-128/Region: 13-residue repeats (G-G-S-S-L-G-Y-G-G-L-Y-G-Y)

RESULT 11
T15126
hypothetical protein T20B6.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000
C;Accession: T15126
R;Beck, C.; Wamsley, P.
submitted to the EMBL Data Library, April 1997
A;Description: The sequence of C. elegans cosmid T20B6.
A;Reference number: 218297
A;Accession: T15126
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-259 <BRC>
A;Cross-references: EMBL:AF000193; NID:g1946986; PID:g1946988; PIDN:AAB52890.1; GSPDB:GN
A;Experimental source: strain Bristol N2; clone T20B6
C;Genetics:
A;Gene: CESP:T20B6.3
A;Introns: 9/2; 231/1
A;Map position: 3
C;Superfamily: Phaseolus glycine-rich protein 1.0

Query Match   62.9%;   Score 109.5; DB 1; Length 155;
Best Local Similarity   76.9%; Pred. No. 7.5e-06;
Matches   30; Conservative 0; Mismatches 0; Indels 9; Gaps   6;

Qy   1 GG--LGYGGL-GYGG--LGYGGL-GYGG--LGYGGL-GY 30
         || |||||| ||| ||| |||| ||| |||| ||
Db   77 GGSSLGYGGLGYGYGGSSLGYGGLYGYGGSSLGYGGLGY 115

RESULT 12
S01421
chorion class B protein (clone M3A5) - silkworm (fragment)

C;Species: Bombyx mori (silkworm)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 20-Aug-1999
C;Accession: S01421
R;Tsitilou, S.G.; Rodakis, G.C.; Alexopoulou, M.; Kafatos, F.C.; Ito, K.; Iatrou, K.
EMBO J. 2, 1845-1852, 1983
A;Title: Structural features of B family chorion sequences in the silkmoth Bombyx mor
A;Reference number: S01420; MUID:84057707
A;Accession: S01421
A;Molecule type: mRNA
A;Residues: 1-91 <TSI>
A;Cross-references: EMBL:X12840; NID:g5816; PIDN:CAA31325.1; PID:g5817
C;Superfamily: chorion class A protein pc292

Query Match   60.1%;   Score 104.5; DB 2; Length 91;
Best Local Similarity   72.7%; Pred. No. 1.6e-05;
Matches   24; Conservative 0; Mismatches 4; Indels 5; Gaps   3;

Qy   1 GGLGYGGLGYG---GLGYGGLGYG-GLGYGGLG 29
         |||||| ||| ||||||| | ||||| |
Db   52 GGLGYGA-GYGGGYGLGYGGYGLGYGGYG 83

RESULT 13
S04513
chorion class B protein (clone M2807) - silkworm (fragment)
C;Species: Bombyx mori (silkworm)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 20-Aug-1999
C;Accession: S04513
R;Tsitilou, S.G.; Rodakis, G.C.; Alexopoulou, M.; Kafatos, F.C.; Ito, K.; Iatrou, K.
EMBO J. 2, 1845-1852 1983
A;Title: Structural features of B family chorion sequences in the silkmoth Bombyx mor
A;Reference number: S01420; MUID:84057707
A;Accession: S04513
A;Molecule type: mRNA
A;Residues: 1-119 <TSI>
A;Cross-references: EMBL:X12838; NID:g5813; PIDN:CAA31323.1; PID:g5812
C;Superfamily: chorion class A protein pc292

Query Match   50.1%;   Score 104.5; DB 2; Length 119;
Best Local Similarity   72.7%; Pred. No. 2e-05;
Matches   24; Conservative 0; Mismatches 4; Indels 5; Gaps   3;

Qy   1 GGLGYGGLGYG---GLGYGGLGYG-GLGYGGLG 29
         |||||| ||| ||||||| | ||||| |
Db   81 GGLGYGA-GYGGGYGLGYGGYGLGYGGYG 112

RESULT 14
T25923
hypothetical protein T27E4.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T25523
R;Bradshaw, H.
submitted to the EMBL Data Library, July 1996
A;Description: The sequence of C. elegans cosmid T27E4.
A;Reference number: 220111
A;Accession: T25923
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-89 <BRA>
A;Cross-references: EMBL:U64837; PIDN:U64837; GSPDB:GN00023; CESP:T27E4.4
A;Experimental source: strain Bristol N2; clone T27E4
C;Genetics:
A;Gene: CESP:T27E4.4
A;Map position: 5
A;Introns: 60/1

Query Match   59.5%;   Score 103.5; DB 2; Length 89;
Best Local Similarity   47.1%; Pred. No. 2e-05;

```
Matches   24; Conservative   0; Mismatches   6; Indels  21; Gaps   3;

Qy     1 GGLGYGGLGYGGL------------GYGGL-GYGGL-GYGGLGY 30
         ||.||||.||||            |||| |||| |||| ||
Db    37 GGYGYGGYGYGGCNCPTYARCTHGPGGYYGYGGYGYGGYYGGYGYGGY 87


RESULT  15
S08381
keratin, 58K type II, cytoskeletal (clone pXenCK55(5/6)) - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 13-Aug-1999
C;Accession: S08381
R;Fouquet, B.; Herrmann, H.; Franz, J.K.; Franke, W.W.
Development 104, 533-548, 1988
A;Title: Expression of intermediate filament proteins during development of Xenopus laev
A;Reference number: S08381; MUID:90032362
A;Accession: S08381
A;Molecule type: mRNA
A;Residues: 1-513 <FOU>
A;Cross-references: EMBL:X14427; NID:g64625; PIDN:CAA32587.1; PID:g64626
C;Superfamily: cytoskeletal keratin


Query Match     58.3%;  Score 101.5; DB 2;  Length 513;
Best Local Similarity   60.0%;  Pred. No. 0.00018;
  Matches  18; Conservative   4; Mismatches   7; Indels   1; Gaps   1;

Qy     1 GGLGYGGLGYGGLGYGGLGY-GGLGYGGLG 29
         ||.||| |:||.||||.|: |: | | |
Db    70 GGAGPGGAGPGGAGYGGSSGFGGGVAYSGPG 99


Search completed: April 24, 2001, 16:42:05
Job time: 470 sec
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.


OM protein - protein search, using sw model

Run on:    April 24, 2001, 16:42:52 ; Search time 44.88 seconds
                                       (without alignments)
                                       22.898 Million cell updates/sec

Title:           US-09-340-736-8
Perfect score:   174
Sequence:        1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        93435 seqs. 34255486 residues

Total number of hits satisfying chosen parameters:    93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No	Score	% Query Match	DB	Length	ID	Description
1	160	92.0	1	119	LAMR_PETMA	petromyzon
2	146	83.9	1	139	LAMP_PETMA	petromyzon
3	140	80.5	1	147	CHB4_BOMMO	bombyx mori
4	136	78.2	1	168	CHB1_BOMMO	bombyx mori
5	122.5	70.1	1	110	CHB6_BOMMO	bombyx mori
6	122	70.1	1	860	ELS_MOUSE	mus musculu
7	122	70.1	1	864	ELS_RAT	rattus norv
8	119	68.4	1	171	CHB2_ANTPO	antheraea p
9	118.5	68.1	1	161	CHB2_BOMMO	bombyx mori
10	115.5	66.4	1	119	CHA1_ANTPO	antheraea p
11	109.5	62.9	1	154	KRSC_CHICK	gallus gall
12	104.5	60.1	1	91	CHB3_BOMMO	bombyx mori
13	104	60.1	1	119	CHB8_BOMMO	bombyx mori
14	102.5	58.9	1	643	K2C1_HUMAN	homo sapien
15	102	58.6	1	152	CU64_LOCMI	locusta mig
16	101.5	58.3	1	512	K2C5_XENLA	xenopus lae
17	101	58.0	1	79	KRHA_RABIT	oryctolagus
18	100	57.5	1	1380	DDX9_MOUSE	mus musculu
19	98	56.3	1	168	GRP2_SORBI	sorghum bic
20	97	55.9	1	126	CHB7_BOMMO	bombyx mori
21	96.5	55.5	1	750	ELS_CHICK	gallus gall
22	95.5	54.9	1	145	CU65_LOCMI	locusta mig
23	94	54.0	1	127	KRCL_CHICK	gallus gall
24	94	54.0	1	129	CHA1_BOMMO	bombyx mori
25	93	53.4	1	486	K1C0_XENLA	xenopus lae
26	93	53.4	1	195	ROAB_ARTSA	artemia sal
27	92	52.9	1	373	RO31_XENLA	xenopus lae
28	92	52.9	1	385	RO32_XENLA	xenopus lae
29	91.5	52.6	1	521	OCIN_MOUSE	mus musculu
30	91	52.3	1	627	K2CL_MOUSE	mus musculu
31	91	52.3	1	157	CU63_LOCMI	locusta mig
32	90.5	51.7	1	169	GRI0_BRANA	brassica na
33	90	51.7	1	208	AC20_TERMO	tenebrio mo
34	89.5	51.4	1	142	GRP1_SORBI	sorghum bic
35	89.5	51.4	1	569	K1C1_MOUSE	mus musculu
36	88	50.6	1	182	K2CA_BOVIN	bos taurus
37	88	50.6	1	747	SPD1_NEPCL	nephila cla
38	87.5	50.3	1	177	CHB1_BOMMO	bombyx mori
39	87	50.0	1	1287	DDX9_BOVIN	bos taurus
40	85	48.9	1	641	EBNI_EBV	epstein-bar
41	84	48.3	1	82	KRHA_SHEEP	ovis aries
42	84	48.3	1	371	ROA1_HUMAN	homo sapien
43	83	47.7	1	130	CHB1_ANTPO	antheraea p
44	83	47.7	1	183	GRP2_ORYSA	oryza sativ
45	83	47.7	1	645	K22E_HUMAN	homo sapien

ALIGNMENTS

RESULT 1
LAMR_PETMA

```
ID   LAMR_PETMA     STANDARD;      PRT;   119 AA.
AC   P35577;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   LAMPRIN 1.8-10 PRECURSOR (CARTILAGE MATRIX PROTEIN).
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 20-44.
RC   TISSUE=Cartilage;
RX   MEDLINE=93123269; PubMed=7678258;
RA   Robson P., Wright G.M., Sitarz E., Maiti A., Rawat M., Youson J.H.,
RA   Keeley F.W.;
RT   "Characterization of lamprin, an unusual matrix protein from lamprey
RT   cartilage. Implications for evolution, structure, and assembly of
RT   elastin and other fibrillar proteins.";
RL   J. Biol. Chem. 268:1440-1447(1993).
CC   -!- FUNCTION: SELF-AGGREGATING PROTEIN THAT MAKES PART OF THE SOLUBLE
CC        FORM OF LAMPRIN.
CC   -!- SUBUNIT: THE POLYMERIC LAMPRIN CHAINS SELF-AGGREGATE TO FORM
CC        FIBERS AND HAVE SECONDARY STRUCTURES PARTICULARLY RICH IN BETA-
CC        SHEETS AND IN BETA-TURNS.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L05926; AAA49270.1; -.
DR   PIR; B45051; B45051.
KW   Cartilage; Repeat; Connective tissue; Signal.
FT   SIGNAL        1     19
FT   CHAIN        20    119       LAMPRIN 1.8-10.
FT   DOMAIN       41     90       7 X 5 AA APPROXIMATE REPEATS.
FT   REPEAT       41     45       1.
FT   REPEAT       46     50       2.
FT   REPEAT       51     55       3.
FT   REPEAT       56     60       4.
FT   REPEAT       61     65       5.
FT   REPEAT       66     70       6.
FT   REPEAT       86     90       7.
SQ   SEQUENCE   119 AA;  11329 MW;  4EB476FACB7591B8 CRC64;

Query Match             92.0%;  Score 160;  DB 1;  Length 119;
Best Local Similarity   93.3%;  Pred. No. 4.6e-10;
Matches   28;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;
```

```
Qy    1  GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
         |||||||||||||||||||||||||| ||||||||
Db   41  GGLGYGGLGYGGLGYGGLGYGGLGVAGLGYGGLGY 70
```

RESULT 2
```
LAMP_PETMA
ID   LAMP_PETMA     STANDARD;     PRT;   139 AA.
AC   P33575; P33576;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   LAMPRIN 0.9 PRECURSOR (CARTILAGE MATRIX PROTEIN).
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 20-44.
RC   TISSUE=Cartilage;
RX   MEDLINE=93123269; PubMed=7678258;
RA   Robson P., Wright G.M., Sitarz E., Maiti A., Rawat M., Youson J.H.,
RA   Keeley F.W.;
RT   "Characterization of lamprin, an unusual matrix protein from lamprey
RT   cartilage. Implications for evolution, structure, and assembly of
RT   elastin and other fibrillar proteins.";
RL   J. Biol. Chem. 268:1440-1447(1993).
CC   -!- FUNCTION: SELF-AGGREGATING PROTEIN THAT MAKES PART OF THE SOLUBLE
CC       FORM OF LAMPRIN
CC   -!- SUBUNIT: THE POLYMERIC LAMPRIN CHAINS SELF-AGGREGATE TO FORM
CC       FIBERS AND HAVE SECONDARY STRUCTURES PARTICULARLY RICH IN BETA-
CC       SHEETS AND IN BETA-TURNS.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC   -!- ALTERNATIVE PRODUCTS: TWO ISOFORM; 0.9-12 (SHOWN HERE) AND 0.9-10;
CC       SEEM TO BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------
DR   EMBL; L05925; AAA49269.1; -.
DR   EMBL; L05924; AAA49268.1; -.
DR   PIR; C45051; C45051.
DR   PIR; A45051; A45051.
KW   Cartilage; Repeat; Connective tissue; Signal; Alternative splicing.
FT   SIGNAL      1     19
FT   CHAIN      20    139       LAMPRIN 0.9.
FT   DOMAIN     42    110       8 X 5 AA APPROXIMATE REPEATS.
FT   REPEAT     42     46       1.
FT   REPEAT     47     51       2.
FT   REPEAT     52     56       3.
FT   REPEAT     57     61       4.
FT   REPEAT     62     66       5.
FT   REPEAT     67     71       6.
FT   REPEAT     92     96       7.
FT   REPEAT    106    110       8.
FT   VARSPLIC   86    104       MISSING (IN ISOFORM 0.9-10).
SQ   SEQUENCE   139 AA;  13257 MW;  E248ABB7A6CEC7C3 CRC64;

Query Match             83.9%;  Score 146;  DB 1;  Length 139;
Best Local Similarity   86.7%;  Pred. No. 1.2e-08;
Matches   26;  Conservative   0;  Mismatches   4;  Indels   0;  Gaps   0;

Qy    1  GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
         |||||||||||||||||||||| ||||||
Db   42  GGLGYGGLGYGGLGVAGLGYGGLGY 71
```

RESULT 3
```
CHB4_BOMMO
ID   CHB4_BOMMO     STANDARD;     PRT;   147 AA.
AC   P05685;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   CHORION CLASS B PROTEIN B.L1 (410) (FRAGMENT).
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=703;
RX   MEDLINE=84057707; PubMed=6571700;
RA   Tsitilou S.G., Rodakis G.C., Alexopoulou M., Kafatos F.C., Ito K.,
RA   Iatrou K.;
RT   "Structural features of B family chorion sequences in the silkmoth
RT   Bombyx mori, and their evolutionary implications.";
RL   EMBO J. 2:1845-1852(1983).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE
CC       SILK MOTH
CC   -!- SIMILARITY: MEMBER OF THE BETA-BRANCH OF CHORION PROTEIN TO WHICH
CC       BELONG CLASSES B, CB AND HCB.
CC   -------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------
DR   EMBL; X01826; CAA25964.1; -.
DR   PIR; S04515; S04515.
DR   InterPro; IPR002635; -.
DR   Pfam; PF01723; Chorion; 1.
KW   Eggshell; Chorion; Repeat; Multigene family.
FT   NON_TER     1      1
FT   DOMAIN      1     38       LEFT ARM.
FT   DOMAIN     39    107       CENTRAL DOMAIN.
FT   DOMAIN    108    147       RIGHT ARM (GLY-RICH TANDEM REPEATS).
FT   DOMAIN     11     30       4 X 5 AA TANDEM REPEATS OF G-Y-G-G-L.
SQ   SEQUENCE   147 AA;  13524 MW;  97DC60A87FF2B04B CRC64;

Query Match             80.5%;  Score 140;  DB 1;  Length 147;
Best Local Similarity   89.3%;  Pred No. 4.9e-08;
Matches   25;  Conservative   0;  Mismatches   3;  Indels   0;  Gaps   0;

Qy    2  GGLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
         | |.|| |||||||||||||||||||||||
Db    4  GRGCGGRGYGGLGYGGLGYGGLGYGGLG 31
```

RESULT 4
```
CHB1_BOMMO
ID   CHB1_BOMMO     STANDARD;     PRT;   168 AA.
AC   P08827;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-NOV-1988 (Rel. 09, Last annotation update)
DE   CHORION CLASS B PROTEIN L11 PRECURSOR.
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=703;
RX   MEDLINE=87060979; PubMed=3023635;
RA   Spoerel N., Nguyen H.T., Kafatos F.C.;
RT   "Gene regulation and evolution in the chorion locus of Bombyx mori.
RT   Structural and developmental characterization of four eggshell genes
RT   and their flanking DNA regions.";
RL   J. Mol. Biol. 190:23-35(1986).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE
CC       SILK MOTH.
CC   -!- SIMILARITY: MEMBER OF THE BETA-BRANCH OF CHORION PROTEIN TO WHICH
CC       BELONG CLASSES B, CB AND HCB.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X15557; CAA33566.1; -.
DR   PIR; C24255; C24255.
DR   InterPro; IPR002635; -.
DR   Pfam; PF01723; Chorion; 1.
KW   Eggshell; Chorion; Repeat; Multigene family; Signal.
FT   SIGNAL        1     21
FT   CHAIN        22    168       CHORION CLASS B PROTEIN L11.
FT   DOMAIN       22     59       LEFT ARM.
FT   DOMAIN       60    128       CENTRAL DOMAIN.
FT   DOMAIN      129    168       RIGHT ARM (GLY-RICH TANDEM REPEATS).
FT   DOMAIN       32     51       4 X 5 AA TANDEM REPEATS OF G-Y-G-G-L.
SQ   SEQUENCE   168 AA;  15594 MW;  BECC357AE942A12B CRC64;

Query Match             78.2%;  Score 136;  DB 1;  Length 168;
Best Local Similarity   85.7%;  Pred. No. 1.3e-07;
Matches   24;  Conservative   0;  Mismatches   4;  Indels   0;  Gaps   0;

Qy          2 GLGYGGLGYGGLGYGGLGYGGLG 29
              |||||||||||||||||||||| |
Db         25 GGGCGGRGYGGLGYGGLGYGGLGYGGYG 52

RESULT 5
CHB6_BOMMO
ID   CHB6_BOMMO     STANDARD;      PRT;   110 AA.
AC   P08917;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-NOV-1988 (Rel. 09, Last annotation update)
DE   CHORION CLASS B PROTEIN M2410 (FRAGMENT).
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=703;
RX   MEDLINE=84057707; PubMed=6571700;
RA   Tsitilou S.G., Rodakis G.C., Alexopoulou M., Kafatos F.C., Ito K.,
RA   Iatrou K.;
RT   "Structural features of B family chorion sequences in the silkmoth
RT   Bombyx mori, and their evolutionary implications.";
RL   EMBO J. 2:1845-1852(1983).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE
CC       SILK MOTH.
CC   -!- SIMILARITY: MEMBER OF THE BETA-BRANCH OF CHORION PROTEIN TO WHICH
CC       BELONG CLASSES B, CB AND HCB.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X12837; CAA31322.1; -.
DR   PIR; S01420; S01420.
DR   InterPro; IPR002635; -.
DR   Pfam; PF01723; Chorion; 1.
KW   Eggshell; Chorion; Repeat; Multigene family.
FT   NON_TER       1      1
FT   DOMAIN        1     27       LEFT ARM.
FT   DOMAIN       28     96       CENTRAL DOMAIN.
FT   DOMAIN       97   >110       RIGHT ARM (GLY-RICH TANDEM REPEATS).
FT   DOMAIN       <1    119       TANDEM REPEATS OF G-Y-G-G-L.
FT   NON_TER     110    110
SQ   SEQUENCE   110 AA;  10243 MW;  412397AA58190BFE CRC64;

Query Match             70.4%;  Score 122.5;  DB 1;  Length 110;
Best Local Similarity   92.0%;  Pred. No. 2e-06;
Matches   23;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   1;

Qy          5 YGGLGYGGLGYGGLGYGGLGYGGLG 29
              |||||||||||||||||||||| || |
Db          1 YGGLGYGGLGYGGLGYGGLG-GGCG 24

RESULT 6
ELS_MOUSE
ID   ELS_MOUSE      STANDARD;      PRT;   860 AA.
AC   P54320;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN).
GN   ELN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Lung;
RX   MEDLINE=95130069; PubMed=7829060;
RA   Wydner K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT   "Use of an intron polymorphism to localize the tropoelastin gene to
RT   mouse chromosome 5 in a region of linkage conservation with human
RT   chromosome 7.";
RL   Genomics 23:125-131(1994).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U08210; AAA80155.1; -.
DR   HSSP; P04002; 1WPA.
DR   MGD; MGI:95317; Eln.
KW   Structural protein; Repeat; Signal; Connective tissue.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    860       ELASTIN.
SQ   SEQUENCE   860 AA;  71955 MW;  0CC0B5AAE1EDD7F1 CRC64;
```

```
Query Match        70.1%;  Score 122;  DB 1;  Length 860;
Best Local Similarity   82.8%;  Pred. No. 1.1e-05;
Matches   24;  Conservative   0;  Mismatches   5;  Indels   0;  Gaps   0;

Qy     1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
         ||| |||| |||| |||| |||| || |||
Db   733 GGLGAGGLGAGGLGAGGLGAGGLGAGGLG 761

RESULT 7
ELS_RAT
ID   ELS_RAT      STANDARD;      PRT;   864 AA.
AC   Q99372;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ELASTIN PRECURSOR (TROPOELASTIN) (FRAGMENT).
GN   ELN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91104868; PubMed=1702999;
RA   Pierce R.A., Deak S.B., Stolle C.A., Boyd C.D.;
RT   "Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";
RL   Biochemistry 29:9677-9683(1990).
RN   [2]
RP   SEQUENCE OF 781-854 FROM N.A.
RX   MEDLINE=88330858; PubMed=2971041;
RA   Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;
RT   "Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.";
RL   J. Biol. Chem. 263:13504-13507(1988).
RN   [3]
RP   SEQUENCE OF 264-533 AND 558-864 FROM N.A.
RX   MEDLINE=92241859; PubMed=1572637;
RA   Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;
RT   "Elements of the rat tropoelastin gene associated with alternative
RT   splicing.";
RL   Genomics 12:651-658(1992).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- ALTERNATIVE PRODUCTS: THREE DIFFERENT MRNAS HAVE BEEN FOUND THAT
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   -------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M60647; AAA2269.1; -.
DR   EMBL; J04035; AAA2268.1; -.
DR   EMBL; M86372; AAA42372.1; -.
DR   EMBL; M86355; AAA42271.1; JOINED.
DR   EMBL; M86363; AAA42271.1; JOINED.
DR   EMBL; M86364; AAA42271.1; JOINED.
DR   EMBL; M86366; AAA42271.1; JOINED.
DR   EMBL; M86371; AAA42271.1; JOINED.
DR   EMBL; M86376; AAA42271.1; JOINED.
DR   EMBL; M86373; AAA42272.1; -.
DR   EMBL; M86375; AAA42272.1; JOINED.
DR   HSSP; P04002; 1WFA.
KW   Structural protein; Connective tissue; Repeat; Signal;
KW   Alternative splicing.
FT   NON_TER       1      1       BY SIMILARITY.
FT   SIGNAL       <1     21       ELASTIN.
FT   CHAIN        22    864       MISSING (IN CERTAIN CLONES).
FT   VARIANT     263    307       MISSING (IN CERTAIN CLONES).
FT   VARIANT     308    308       MISSING (IN CERTAIN CLONES).
FT   VARIANT     809    823       MISSING (IN CERTAIN CLONES).
SQ   SEQUENCE   864 AA;  72786 MW;  456894B09E79FD4 CRC64;

Query Match        70.1%;  Score 122;  DB 1;  Length 864;
Best Local Similarity   82.8%;  Pred. No. 1.1e-05;
Matches   24;  Conservative   0;  Mismatches   5;  Indels   0;  Gaps   0;

Qy     1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
         ||| |||| |||| |||| |||| || |||
Db   738 GGLGAGGLGAGGLGAGGLGAGGLGAGGLG 766

RESULT 8
CHB4_ANTPO
ID   CHB4_ANTPO      STANDARD;      PRT;   171 AA.
AC   P02847;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CHORION CLASS B PROTEIN PC401 PRECURSOR (FRAGMENT).
OS   Antheraea polyphemus (Polyphemus moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Saturniidae; Saturniinae; Antheraea.
OX   NCBI_TaxID=7120;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=80090072; PubMed=519771;
RA   Jones C.W., Rosenthal N., Rodakis G.C., Kafatos F.C.;
RT   "Evolution of two major chorion multigene families as inferred from
RT   cloned cDNA and protein sequences.";
RL   Cell 18:1317-1332(1979).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE
CC       SILK MOTH.
CC   -!- SIMILARITY: MEMBER OF THE BETA-BRANCH OF CHORION PROTEIN TO WHICH
CC       BELONG CLASSES B, CB AND HCB.
CC   -------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; V00077; CAA23419.1; -.
DR   PIR; A03337; JBQ41.
DR   InterPro; IPR002635; -.
DR   Pfam; PF01723; Chorion; 1.
KW   Eggshell; Chorion; Repeat; Multigene family; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1     18
FT   CHAIN        19    171       CHORION CLASS B PROTEIN PC401.
FT   DOMAIN       19     61       LEFT ARM.
FT   DOMAIN       62    132       CENTRAL DOMAIN.
FT   DOMAIN      133    171       RIGHT ARM (GLY-RICH TANDEM REPEATS).
SQ   SEQUENCE   171 AA;  16081 MW;  6D38290D4EE5692E CRC64;

Query Match        68.4%;  Score 119;  DB 1;  Length 171;
Best Local Similarity   75.9%;  Pred. No. 6.2e-05;
Matches   22;  Conservative   1;  Mismatches   6;  Indels   0;  Gaps   0;

Qy     1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
         |||||||||| |:||||||| |:|| ||
Db   136 GGLGYGGLGYGGVGGYGLGYGYGLGGCG 164
```

RESULT 9
CHB2_BOMMO
ID   CHB2_BOMMO   STANDARD;   PRT;   161 AA.
AC   P08828;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-NOV-1988 (Rel. 09, Last annotation update)
DE   CHORION CLASS B PROTEIN L12 PRECURSOR.
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87060979; PubMed=3023635;
RA   Spoerel N., Nguyen H.T., Kafatos F.C.;
RT   "Gene regulation and evolution in the chorion locus of Bombyx mori.
RT   Structural and developmental characterization of four eggshell genes
RT   and their flanking DNA regions.";
RL   J. Mol. Biol. 190:23-35(1986).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE
CC   SILK MOTH.
CC   -!- SIMILARITY: MEMBER OF THE BETA-BRANCH OF CHORION PROTEIN TO WHICH
CC   BELONG CLASSES B, CB AND BCB.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X15557; CAA33564.1; -.
DR   PIR; D24255; D24255.
DR   InterPro; IPR002635; -.
DR   Pfam; PF01723; Chorion; 1.
KW   Eggshell; Chorion; Repeat; Multigene family; Signal.
FT   SIGNAL   1   21
FT   CHAIN   22   161   CHORION CLASS B PROTEIN L12.
FT   DOMAIN   22   52   LEFT ARM.
FT   DOMAIN   53   121   CENTRAL DOMAIN.
FT   DOMAIN   122   161   RIGHT ARM (GLY-RICH TANDEM REPEATS).
FT   DOMAIN   30   44   3 X 5 AA TANDEM REPEATS OF G-Y-G-G-L.
SQ   SEQUENCE   161 AA; 15123 MW; 782FDC49FC07FA0B CRC64;

Query Match   68.8%; Score 118.5; DB 1; Length 161;
Best Local Similarity   82.1%; Pred. No. 6.6e-06;
Matches   23; Conservative   0; Mismatches   4; Indels   1; Gaps   1;

Qy   2 GLGYGGLGYGGLGYGGLGYGGLGYGGLG 29
          | | || ||||||||||||| || ||
Db   23 GCGCGGRGYGGLYGGLGYGGLG-GGCG 49

RESULT 10
CHA1_ANTPO
ID   CHA1_ANTPO   STANDARD;   PRT;   119 AA.
AC   P02846;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-NOV-1988 (Rel. 09, Last annotation update)
DE   CHORION CLASS A PROTEIN PC292 PRECURSOR (FRAGMENT).
OS   Antheraea polyphemus (Polyphemus moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Saturniidae; Saturniinae; Antheraea.
OX   NCBI_TaxID=7120;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81053859; PubMed=7433133;
RA   Tsitilou S.G., Regier J.C., Kafatos F.C.;
RT   "Selection and sequence analysis of a cDNA clone encoding a known
RT   chorion protein of the A family.";
RL   Nucleic Acids Res. 8:1967-1997(1980).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE
CC   SILK MOTH.
CC   -!- SIMILARITY: MEMBER OF THE ALPHA-BRANCH OF CHORION PROTEIN TO WHICH
CC   BELONG CLASSES A, CA AND HCA.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; J01159; AAA27780.1; -.
DR   EMBL; V00078; CAA23420.1; -.
DR   PIR; A03336; JRAO92.
KW   Eggshell; Chorion; Repeat; Multigene family; Signal.
FT   NON_TER   1   1
FT   SIGNAL   <1   6
FT   CHAIN   7   119   CHORION CLASS A PROTEIN PC292.
FT   DOMAIN   7   53   LEFT ARM.
FT   DOMAIN   54   102   CENTRAL DOMAIN.
FT   DOMAIN   103   119   RIGHT ARM.
FT   REPEAT   28   32
FT   REPEAT   33   37
FT   REPEAT   38   42
FT   REPEAT   43   47
FT   VARIANT   118   118   P -> L (PROBABLE POLYMORPHISM).
SQ   SEQUENCE   119 AA; 11091 MW; 65E01B7C0F2A2A4A CRC64;

Query Match   66.4%; Score 115.5; DB 1; Length 119;
Best Local Similarity   58.6%; Pred. No. 1e-05;
Matches   24; Conservative   0; Mismatches   6; Indels   5; Gaps   1;

Qy   1 GGLGYGGL----GYGGLGYGGLGYGGLGYGGLGY 30
          | ||| || ||||| ||||| || || |||
Db   10 GGLGLKGLAAPACGCGGLGYGGLGYGALGYGGLGY 44

RESULT 11
KRSC_CHICK
ID   KRSC_CHICK   STANDARD;   PRT;   154 AA.
AC   P04459;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-MAY-1992 (Rel. 22, Last annotation update)
DE   KERATIN, SCALE (S-KER).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84158528; PubMed=6200321;
RA   Gregg K., Wilton S.D., Parry D.A.D., Rogers G.E.;
RT   "A comparison of genomic coding sequences for feather and scale
RT   keratins: structural and evolutionary implications.";
RL   EMBO J. 3:175-178(1984).
CC   -!- MISCELLANEOUS: THE AVIAN KERATINS (F-KER, S-KER, C-KER AND B-KER)
CC   ARE A COMPLEX MIXTURE OF VERY SIMILAR POLYPEPTIDES.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way

```
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; X00315; CAA25084.1; ALT_SEQ.
DR   PIR; A02851; KRCHS.
KW   Keratin; Fibrous protein; Multigene family; Repeat.
FT   INIT_MET       0      0                   BY SIMILARITY.
FT   REPEAT        69     75
FT   REPEAT        76     88
FT   REPEAT        89    101
FT   REPEAT       102    114
FT   REPEAT       115    127
SQ   SEQUENCE     154 AA;  15710 MW;  6SB337485EC00466 CRC64;

Query Match             62.9%;  Score 109.5;  DB 1;  Length 154;
Best Local Similarity   76.9%;  Pred. No. 4.8e-05;
Matches   30;  Conservative    0;  Mismatches    0;  Indels    9;  Gaps    6;

Qy      1 GG--LGYGGL-GYGG--LGYGGL-GYGG--LGYGGL-GY 30
          ||  ||||| |||  ||||| |||  ||||| |||||| ||
Db     76 GGSSLGYGGLXYGGSSLGYGGLXYGGSSLGYGGSLGYGGLGY 114

RESULT 12
CHB3_BOMMO
ID   CHB3_BOMMO      STANDARD;      PRT;    91 AA.
AC   P08915;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-NOV-1988 (Rel. 09, Last annotation update)
DE   CHORION CLASS B PROTEIN M3A5 (FRAGMENT).
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=703;
RA   Tsitilou S.G., Rodakis G.C., Alexopoulou M., Kafatos F.C., Ito K.,
RA   Iatrou K.;
RT   "Structural features of B family chorion sequences in the silkmoth
RT   Bombyx mori, and their evolutionary implications.";
RL   EMBO J. 2:1845-1852(1983).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE
CC       SILK MOTH.
CC   -!- SIMILARITY: MEMBER OF THE BETA-BRANCH OF CHORION PROTEIN TO WHICH
CC       BELONG CLASSES B, CB AND HCB.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; X12840; CAA31325.1; -.
DR   PIR; S01421; S01421.
DR   InterPro; IPR002635; -.
DR   Pfam; PF01723; Chorion; 1.
KW   Eggshell; Chorion; Repeat; Multigene family.
FT   NON_TER        1      1
FT   DOMAIN        <1     51                   CENTRAL DOMAIN.
FT   DOMAIN        52     91                   RIGHT ARM (GLY-RICH TANDEM REPEATS).
SQ   SEQUENCE      91 AA;   8645 MW;  7A72057AB78460F3 CRC64;

Query Match             60.1%;  Score 104.5;  DB 1;  Length 91;
Best Local Similarity   72.7%;  Pred. No. 9.7e-05;
```

```
Matches   24;  Conservative    0;  Mismatches    4;  Indels    5;  Gaps    3;

Qy      1 GGLGYGGLGYG---LGYGGLCYG-GLGYGGLG 29
          ||||| ||     ||||||||| |||||||| |
Db     52 GGLGYGA-GYGGGYGLGYGGGYGLGYGGYG 83

RESULT 13
CHB8_BOMMO
ID   CHB8_BOMMO      STANDARD;      PRT;    119 AA.
AC   P08911;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-NOV-1988 (Rel. 09, Last annotation update)
DE   CHORION CLASS B PROTEIN M2807 (FRAGMENT).
OS   Bombyx mori (Silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Bombycoidea; Bombycidae; Bombyx.
OX   NCBI_TaxID=7091;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=703;
RX   MEDLINE=84057707; PubMed=6571700;
RA   Tsitilou S.G., Rodakis G.C., Alexopoulou M., Kafatos F.C., Ito K.,
RA   Iatrou K.;
RT   "Structural features of B family chorion sequences in the silkmoth
RT   Bombyx mori, and their evolutionary implications.";
RL   EMBO J. 2:1845-1852(1983).
CC   -!- FUNCTION: THIS PROTEIN IS ONE OF MANY FROM THE EGGSHELL OF THE
CC       SILK MOTH.
CC   -!- SIMILARITY: MEMBER OF THE BETA-BRANCH OF CHORION PROTEIN TO WHICH
CC       BELONG CLASSES B, CB AND HCB.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; X12838; CAA31323.1; -.
DR   PIR; S04513; S04513.
DR   InterPro; IPR002635; -.
DR   Pfam; PF01723; Chorion; 1.
KW   Eggshell; Chorion; Repeat; Multigene family.
FT   NON_TER        1      1
FT   DOMAIN        <1     11                   LEFT ARM.
FT   DOMAIN        12     80                   CENTRAL DOMAIN.
FT   DOMAIN        81   >119                   RIGHT ARM (GLY-RICH TANDEM REPEATS).
FT   NON_TER      119    119
SQ   SEQUENCE     119 AA;  10957 MW;  8B26AA85F7EPAD28 CRC64;

Query Match             60.1%;  Score 104.5;  DB 1;  Length 119;
Best Local Similarity   72.7%;  Pred. No. 0.00012;
Matches   24;  Conservative    0;  Mismatches    4;  Indels    5;  Gaps    3;

Qy      1 GGLGYGGLGYG---LGYGGLGYG-GLGYGGLG 29
          ||||| ||     ||||||||| |||||||| |
Db     81 GGLGYGA-GYGDSYGLGYGGGYGLGYGGGYG 112

RESULT 14
K2C1_HUMAN
ID   K2C1_HUMAN      STANDARD;      PRT;    643 AA.
AC   P04264;
DT   20-MAR-1987 (Rel. 04, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   KERATIN, TYPE II CYTOSKELETAL 1 (CYTOKERATIN 1) (K1) (CK 1) (67 KDA
DE   CYTOKERATIN) (HAIR ALPHA PROTEIN).
```

```
GN   KRT1 OR KRTA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85166239; PubMed=2580302;
RA   Johnson L.D., Idler W.W., Zhou X.-M., Roop D.R., Steinert P.M.;
RT   "Structure of a gene for the human epidermal 67-kDa keratin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:1896-1900(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85207740; PubMed=2581964;
RA   Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,
RA   Roop D.R.;
RT   "Amino acid sequences of mouse and human epidermal type II keratins
RT   of Mr 67,000 provide a systematic basis for the structural and
RT   functional diversity of the end domains of keratin intermediate
RT   filament subunits.";
RL   J. Biol. Chem. 260:7142-7149(1985).
RN   [3]
RP   PRELIMINARY SEQUENCE OF 151-643 FROM N.A.
RA   Whittock N.V., Eady R.A., McGrath J.A.;
RT   "Genomic organization of the human keratin 1 gene.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   REVISIONS, AND VARIANT EHK PRO-160.
RX   MEDLINE=92386601; PubMed=1381288;
RA   Chipev C.C., Korge B.P., Markova N., Bale S.J., Digiovanna J.J.,
RA   Compton J.G., Steinert P.M.;
RT   "A leucine-->proline mutation in the H1 subdomain of keratin 1
RT   causes epidermolytic hyperkeratosis.";
RL   Cell 70:821-828(1992).
RN   [5]
RP   VARIANTS EHK GLY-154; SER-187 AND PRO-192.
RX   MEDLINE=94117869; PubMed=7507151;
RA   Yang J.-M., Chipev C.C., Digiovanna J.J., Bale S.J., Marekov L.N.,
RA   Steinert P.M., Compton J.G.;
RT   "Mutations in the H1 and 1A domains in the keratin 1 gene in
RT   epidermolytic hyperkeratosis.";
RL   J. Invest. Dermatol. 102:17-23(1994).
RN   [6]
RP   VARIANTS EHK PRO-185 AND SER-187.
RX   MEDLINE=94117870; PubMed=7507152;
RA   McLean W.H.I., Eady R.A., Dopping-Hepenstal P.J., McMillan J.R.,
RA   Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,
RA   Morley S.M.;
RT   "Mutations in the rod 1A domain of keratins 1 and 10 in bullous
RT   congenital ichthyosiform erythroderma (BCIE).";
RL   J. Invest. Dermatol. 102:24-30(1994).
RN   [7]
RP   VARIANT EHK GLN-489.
RX   MEDLINE=92376531; PubMed=1380725;
RA   Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,
RA   Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Roop D.R.;
RT   "Mutations in the rod domains of keratins 1 and 10 in epidermolytic
RT   hyperkeratosis.";
RL   Science 257:1128-1130(1992).
RN   [8]
RP   VARIANT ALLELE 1B.
RX   MEDLINE=93107743; PubMed=1281859;
RA   Korge B.P., Compton J.G., Steinert P.M., Mischke D.;
RT   "The two size alleles of human keratin 1 are due to a deletion in the
RT   glycine-rich carboxyl-terminal V2 subdomain.";
RL   J. Invest. Dermatol. 99:697-702(1992).
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC       KERATIN 1 IS GENERALLY ASSOCIATED WITH KERATIN 10.
CC   -!- TISSUE SPECIFICITY: THE SOURCE OF THIS PROTEIN IS NEONATAL
CC       FORESKIN. THE 67-KDA TYPE II KERATINS ARE EXPRESSED IN TERMINALLY
CC       DIFFERENTIATING EPIDERMIS.
CC   -!- POLYMORPHISM: THERE ARE TWO SIZE VARIANTS OF KRT1, TERMED 1A AND
CC       1B WITH ALLELIC FREQUENCIES OF 0.61 AND 0.39. 1B LACKS 7 LACKS
CC       7 RESIDUES COMPARED TO 1A.
CC   -!- DISEASE: DEFECTS IN KRT10 AND KRT1 ARE THE CAUSE OF EPIDERMOLYTIC
CC       HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL
CC       ICHTHYOSIFORM ERYTHRODERMA (BIE)); A HEREDITARY SKIN DISORDER
CC       CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM
CC       CORNEUM.
CC   -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC       MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 KDA) [K9 TO K20] AND II
CC       (NEUTRAL TO BASIC; 56-70 KDA) [K1 TO K8].
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
```

```
DR   EMBL; AF237621; AAF60327.1; -.
DR   EMBL; M98776; AAB47721.1; -.
DR   PIR; A02950; KRHU2.
DR   PIR; A22940; A22940.
DR   HSSP; P05412; 1POS.
DR   Aarhus/Ghent-2DPAGE; 4606; NEPHGE.
DR   MIM; 113800; -.
DR   MIM; 139350; -.
DR   InterPro; IPR001664; -.
DR   InterPro; IPR003054; -.
DR   Pfam; PF00038; filament; 1.
DR   PRINTS; PR01276; TYPE2KERATIN.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin;
KW   Disease mutation; Polymorphism; Phosphorylation.
FT   INIT_MET       0      0
FT   DOMAIN         1    178       HEAD.
FT   DOMAIN       179    488       ROD.
FT   DOMAIN       489    643       TAIL.
FT   DOMAIN       179    214       COIL 1A.
FT   DOMAIN       215    233       LINKER 1.
FT   DOMAIN       234    325       COIL 1B.
FT   DOMAIN       326    349       LINKER 12.
FT   DOMAIN       350    488       COIL 2.
FT   SITE         432    432       STUTTER.
FT   DOMAIN         1    150       GLY/PHE/SER-RICH.
FT   DOMAIN       501    640       GLY/SER-RICH.
FT   MOD_RES       65     65       PHOSPHORYLATION (BY SIMILARITY).
FT   VARIANT      154    154       V -> G (IN EHK).
FT                                 /FTId=VAR_003853.
FT   VARIANT      160    160       L -> P (IN EHK).
FT                                 /FTId=VAR_003854.
FT   VARIANT      185    185       S -> P (IN EHK).
FT                                 /FTId=VAR_003855.
FT   VARIANT      187    187       N -> H (IN EHK).
FT                                 /FTId=VAR_003856.
FT   VARIANT      192    192       S -> P (IN EHK).
FT                                 /FTId=VAR_003857.
FT   VARIANT      311    311       I -> V.
FT                                 /FTId=VAR_003858.
FT   VARIANT      329    329       I -> T.
FT                                 /FTId=VAR_003859.
FT   VARIANT      357    357       N -> Y.
FT                                 /FTId=VAR_003860.
FT   VARIANT      489    489       E -> Q (IN EHK).
FT                                 /FTId=VAR_003861.
FT   VARIANT      536    536       G -> C.
FT                                 /FTId=VAR_003862.
FT   VARIANT      632    632       R -> K.
FT                                 /FTId=VAR_003863.
FT   VARIANT      559    565       MISSING (IN ALLELE 1B).
FT                                 /FTId=VAR_003864.
SQ   SEQUENCE   643 AA;   65886 MW;  DF945DC46225785D CRC64;
```

Job time: 459 sec

```
Query Match            58.9%;  Score 102.5; DB 1;  Length 643;
Best Local Similarity  66.7%;  Pred. No. 0.00072;
Matches  20; Conservative  3; Mismatches  6;  Indels  1;  Gaps  1;

Qy    1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
          || |:|| :|| :|| :|:| |:|| |||| |
Db  114 GGGGFGGFTSGGGFGGGFGGGFGGGTGG-GY 142


RESULT 15
ID  CU64_LOCMI      STANDARD;        PRT;      152 AA.
AC  P11739;
DT  01-OCT-1989 (Rel. 12, Created)
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  01-FEB-1996 (Rel. 33, Last annotation update)
DE  CUTICLE PROTEIN 64 (LM-64) (LM-ACP 64).
OS  Locusta migratoria (Migratory locust).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC  Acridomorpha; Acridoidea; Acrididae; Locusta.
OC  NCBI_TaxID=7004;
RN  [1]
RP  SEQUENCE.
RC  TISSUE=Wing;
RX  MEDLINE=95291280; PubMed=7773253;
RA  Krogh T.N., Skou L., Roepstorff P., Andersen S.O., Hoejrup P.;
RT  "Primary structure of proteins from the wing cuticle of the migratory
RT  locust, Locusta migratoria.";
RL  Insect Biochem. Mol. Biol. 25:319-329(1995).
RN  [2]
RP  SEQUENCE OF 1-59.
RX  MEDLINE=86108304; PubMed=3943519;
RA  Hoejrup P., Andersen S.O., Roepstorff P.;
RT  "Isolation, characterization, and N-terminal sequence studies of
RT  cuticular proteins from the migratory locust, Locusta migratoria.";
RL  Eur. J. Biochem. 154:153-159(1986).
CC  -!- FUNCTION: COMPONENT OF THE CUTICLE OF MIGRATORY LOCUST WHICH
CC      CONTAINS MORE THAN 100 DIFFERENT STRUCTURAL PROTEINS.
CC  -!- DOMAIN: THE TETRAPEPTIDE (A-A-P-(AV)) REPEATS FOUND THROUGHOUT THE
CC      PROTEIN ARE ALSO PRESENT IN MANY PROTEINS CONSTITUTING THE
CC      PROTECTIVE ENVELOPE OF OTHER SPECIES.
CC  -!- MASS SPECTROMETRY: MW=14050.8; METHOD=ELECTROSPRAY.
DR  PIR; H24802; H24802.
KW  Structural protein; Cuticle; Repeat.
FT  REPEAT        27     30        1.
FT  REPEAT        33     37        2.
FT  REPEAT        39     42        3.
FT  REPEAT        86     89        4.
FT  REPEAT        92     95        5.
FT  REPEAT        98    101        6.
FT  REPEAT       127    130        7.
FT  CONFLICT      11     11        G -> S (IN REF. 2).
FT  CONFLICT      21     22        HG -> PH (IN REF. 2).
FT  CONFLICT      27     27        A -> Y (IN REF. 2).
FT  CONFLICT      30     32        AAI -> LAL (IN REF. 2).
FT  CONFLICT      34     34        A -> Y (IN REF. 2).
FT  CONFLICT      45     45        H -> Q (IN REF. 2).
SQ  SEQUENCE    152 AA;  14050 MW;  0C127C22F28E2941 CRC64;


Query Match            58.6%;  Score 102; DB 1;  Length 152;
Best Local Similarity  55.3%;  Pred. No. 0.00026;
Matches  26; Conservative  0; Mismatches  3;  Indels  18;  Gaps  4;

Qy    2 GLGYGG------LGYG---GLGYG----GLGYG---GLGYGGLGY 30
          ||||||      |||   ||||        ||||| ||||:| ||
Db  103 GLGYGGYGHGLSLGYGHGLGLGYAAAPALSLGYGGYGHGLGYGYGYGY 149
```

Search completed: April 24, 2001, 16:42:53

us-09-340-736-8.rsp

```
                GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        April 24, 2001, 16:40:39 ; Search time 125.5 seconds
                                           (without alignments)
                                           28.018 Million cell updates/sec

Title:          US-09-340-736-8
Perfect score:  174
Sequence:       1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

?rched:         374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters:    374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SPTREMBL_15:*
                1: sp_archea:*
                2: sp_bacteria:*
                3: sp_fungi:*
                4: sp_human:*
                5: sp_invertebrate:*
                6: sp_mammal:*
                7: sp_mhc:*
                8: sp_organelle:*
                9: sp_phage:*
               10: sp_plant:*
               11: sp_rodent:*
               12: sp_unclassified:*
               13: sp_vertebrate:*
               14: sp_virus:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
                                SUMMARIES

Result          %       Query
No.     Score   Match   Length  DB  ID      Description
-----   -----   -----   ------  --  ------  ------------------
  1     160     92.0    138     13  Q9YHE0  Q9yhe0 petromyzon
  2     146     83.9    119     13  Q9YGD0  Q9ygd0 petromyzon
  3     146     83.9    138     13  Q9PSW3  Q9psw3 petromyzon
  4     146     83.9    138     13  Q9PSW2  Q9psw2 petromyzon
  5     120     69.0    142     5   Q9U517  Q9u517 manduca sex
  6     116     66.7    33      5   Q17071  Q17071 antheraea p
  7     116     66.7    62      13  P82170  P82170 locusta mig
  8     109.5   62.9    66      13  P87483  P87483 gallus gal
  9     109.5   62.9    904     5   O76271  O76271 mytilus edu
 10     108     62.1    181     5   Q9VMK5  Q9vmk5 drosophila
 11     106.5   61.2    259     5   O02049  O02049 caenorhabdi
 12     105.5   60.6    77      11  O08631  O08631 mus musculu
 13     105.5   60.6    78      11  O09048  O09048 mus musculu
 14     104     59.8    54      11  O08636  O08636 mus musculu
 15     103.5   59.5    89      5   Q23052  Q23052 caenorhabdi
 16     103.5   59.5    287     5   Q17200  Q17200 bombyx mori
 17     103.5   59.5    303     5   Q17201  Q17201 bombyx mori
 18     102     58.6    945     5   Q9W4Y1  Q9w4y1 drosophila
 19     102     58.6    1884    5   Q9NHW2  Q9nhw2 nephila mad
```

```
 20     98.5    56.6    32      5   Q17070  Q17070 antheraea p
 21     98      56.3    181     5   Q9VRM4  Q9vrm4 drosophila
 22     98      56.3    783     2   Q9XAI1  Q9xai1 streptomyce
 23     97.5    56.0    141     5   Q9VNN3  Q9vnn3 drosophila
 24     97      55.7    94      5   Q9YH54  Q9yh54 gallus gall
 25     97      55.7    121     5   Q20726  Q20726 caenorhabdi
 26     96.5    55.5    496     13  Q9V5U8  Q9v5u8 drosophila
 27     96.5    55.5    131     13  O42434  O42434 notophthalm
 28     96      55.2    136     5   O44356  O44356 aequipecten
 29     96      55.2    62      11  O44354  O44354 aequipecten
 30     95.5    54.9    113     11  O08632  O08632 mus musculu
 31     95      54.6    440     5   Q9V5U5  Q9v5u5 drosophila
 32     95      54.6    626     5   Q9V6M1  Q9v6m1 drosophila
 33     95      54.6    102     5   Q9NHW1  Q9nhw1 nephila mad
 34     94.5    54.3    165     5   O19176  O19176 caenorhabdi
 35     94      54.0    2249    10  Q40425  Q40425 nicotiana s
 36     94      54.0    146     5   Q9NHW4  Q9nhw4 nephila cla
 37     93      53.4    158     5   Q9XYI1  Q9xyi1 ciona intes
 38     93      53.4    462     5   Q9NHW3  Q9nhw3 nephila cla
 39     93      53.4    871     5   O44358  O44358 nephila cla
 40     93      53.4    907     5   O44359  O44359 nephila cla
 41     93      53.4    966     3   Q01385  Q01385 neurospora
 42     92      52.6    680     10  P93658  P93658 brassica na
 43     91.5    52.6    49      10  P93517  P93517 zea mays (m
 44     91.5    52.6    341     2   Q9RTY3  Q9rty3 deinococcus
 45
```

                                ALIGNMENTS

```
RESULT 1
Q9YHE0
ID      Q9YHE0        PRELIMINARY;       PRT;    138 AA.
AC      Q9YHE0;
DT      01-MAY-1999 (TrEMBLrel. 10, Created)
DT      01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT      01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE      LAMPRIN L-1.8-12.
GN      L-1.8.
OS      Petromyzon marinus (Sea lamprey).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hypercartia;
OC      Petromyzontiformes; Petromyzontidae; Petromyzon.
OX      NCBI_TaxID=7757;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93123269; PubMed=7678258;
RA      Robson P., Wright G.M., Sitarz E., Maiti A., Rawat M., Youson J.H.,
RA      Keeley F.W.;
RT      "Characterization of lamprin, an unusual matrix protein from lamprey
RT      cartilage. Implications for evolution, structure, and assembly of
RT      elastin and other fibrillar proteins.";
RL      J. Biol. Chem. 268:1440-1447(1993).
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Robson P., Youson J.H., Keeley F.W.;
RL      Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AF097763; AAC97503.1; -.
DR      EMBL; AF097760; AAC97760.1; JOINED.
DR      EMBL; AF097761; AAC97503.1; JOINED.
DR      EMBL; AF097762; AAC97503.1; JOINED.
SQ      SEQUENCE 138 AA;  13328 MW;  EA6D4FE317197C88 CRC64.

Query Match           92.04;   Score 160; DB 13; Length 138;
Best Local Similarity 93.3%;   Pred. No. 2.3e-11;
Matches  28; Conservative  0; Mismatches  2; Indels  0; Gaps  0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
         ||||||||||||||||||||||||| |||||
Do  41 GGLGYGGLGYGGLGYGGLGLGVAGLGYGGLGY 70
```

RESULT 2
ID   Q9YGD0
AC   Q9YGD0;
ID   Q9YGD0   PRELIMINARY;   PRT;   119 AA.
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   LAMPRIN L-0.9-10.
GN   L-0.9.
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93123269; PubMed=7678258;
RA   Robson P., Wright G.M., Sitarz E., Maiti A., Rawat M., Youson J.H.,
RA   Keeley F.W.;
RT   "Characterization of lamprin, an unusual matrix protein from lamprey
RT   cartilage. Implications for evolution, structure, and assembly of
RT   elastin and other fibrillar proteins.";
RL   J. Biol. Chem. 268:1440-1447(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Robson P., Youson J.H., Keeley F.W.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF097759; AAC97500.1; -.
DR   EMBL; AF097758; AAC97500.1; JOINED.
DR   EMBL; AF097757; AAC97498.1; -.
DR   EMBL; AF097756; AAC97498.1; JOINED.
SQ   SEQUENCE 119 AA; 11145 MW; BB608EE263DF03E0 CRC64;

Query Match   83.9%; Score 146; DB 13; Length 119;
Best Local Similarity   86.7%;
  Matches   26; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
              ||||||||||||||| ||| ||||||||||
Db   41 GGLGYGGLGYGGLGYGGLGVAGLGYGGLGY 70

RESULT 3
ID   Q9PSW3
AC   Q9PSW3;
ID   Q9PSW3   PRELIMINARY;   PRT;   138 AA.
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LAMPRIN L-0.8-12.
GN   L-0.8.
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93123269; PubMed=7678258;
RA   Robson P., Wright G.M., Sitarz E., Maiti A., Rawat M., Youson J.H.,
RA   Keeley F.W.;
RT   "Characterization of lamprin, an unusual matrix protein from lamprey
RT   cartilage. Implications for evolution, structure, and assembly of
RT   elastin and other fibrillar proteins.";
RL   J. Biol. Chem. 268:1440-1447(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Robson P., Youson J.H., Keeley F.W.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF097757; AAC97499.1; -.
DR   EMBL; AF097756; AAC97499.1; JOINED.
SQ   SEQUENCE 138 AA; 13094 MW; AB2F35D0D25D7580 CRC64;

Query Match   83.9%; Score 146; DB 13; Length 138;
Best Local Similarity   86.7%;   Pred. No. 8.1e-10;
  Matches   26; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
              ||||||||||||||| ||| ||||||||||
Db   41 GGLGYGGLGYGGLGVAGLGYGGLGY 70

RESULT 4
ID   Q9PSW2
AC   Q9PSW2;
ID   Q9PSW2   PRELIMINARY;   PRT;   138 AA.
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   LAMPRIN L-0.9.
GN   L-0.9.
OS   Petromyzon marinus (Sea lamprey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC   Petromyzontiformes; Petromyzontidae; Petromyzon.
OX   NCBI_TaxID=7757;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93123269; PubMed=7678258;
RA   Robson P., Wright G.M., Sitarz E., Maiti A., Rawat M., Youson J.H.,
RA   Keeley F.W.;
RT   "Characterization of lamprin, an unusual matrix protein from lamprey
RT   cartilage. Implications for evolution, structure, and assembly of
RT   elastin and other fibrillar proteins.";
RL   J. Biol. Chem. 268:1440-1447(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Robson P., Youson J.H., Keeley F.W.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF097759; AAC97501.1; -.
DR   EMBL; AF097758; AAC97501.1; JOINED.
SQ   SEQUENCE 138 AA; 13144 MW; AB2F35D0C74924D0 CRC64;

Query Match   83.9%; Score 146; DB 13; Length 138;
Best Local Similarity   86.7%;   Pred. No. 8.1e-10;
  Matches   26; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
              ||||||||||||||| ||| ||||||||||
Db   41 GGLGYGGLGYGGLGVAGLGYGGLGY 70

RESULT 5
ID   Q9U517
AC   Q9U517;
ID   Q9U517   PRELIMINARY;   PRT;   142 AA.
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PUTATIVE CUTICLE PROTEIN.
OS   Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Sphingoidea; Sphingidae; Sphinginae; Manduca.
OX   NCBI_TaxID=7130;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Robertson H.M., Martos R., Sears C.R., Todres E.Z., Walden K.K.O.,
RA   Nardi J.B.;
RT   "Diversity of odourant binding proteins revealed by an expressed
RT   sequence tag project on male Manduca sexta moth antennae.";
RL   Insect Mol. Biol. 8:501-518(1999).
DR   EMBL; AF117571; AAF16693.1; -.
SQ   SEQUENCE 142 AA; 14112 MW; 17230DE0554DEE0 CRC64;

Query Match   83.9%; Score 146; DB 13; Length 138;
Best Local Similarity   86.7%;   Pred. No. 8.1e-10;
  Matches   26; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
              ||||||||||||||| ||| ||||||||||
Db   41 GGLGYGGLGYGGLGVAGLGYGGLGY 70

Query Match　　　69.0%;　Score 120;　DB 5;　Length 142;
Best Local Similarity　77.4%;　Pred. No. 6.5e-07;
Matches　24;　Conservative　1;　Mismatches　2;　Indels　4;　Gaps　2;

Qy　　2 GLGYGGLGYG--GLGYGGSLGYGSLGYGGLGY 30
　　　　 |||||||||| ||:|| |||| ||||
Db　　29 GLGYGGLGYAGHGVGYDGLGYG--GYGGLGY 57

RESULT　6
ID　Q17071
AC　Q17071;
DT　01-NOV-1996 (TrEMBLrel. 01, Created)
DT　01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT　01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE　SILKMOTH (A.POLYPHEMUS) CHORION PROTEIN (FRAGMENT).
OS　Antheraea polyphemus (Polyphemus moth).
OC　Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC　Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC　Bombycoidea; Saturniidae; Saturniinae; Antheraea.
OX　NCBI_TaxID=7120;
RN　[1]
RP　SEQUENCE FROM N.A.
RX　MEDLINE=80090072; PubMed=519771;
RA　Jones C.W., Rosenthal N., Rodakis G.C., Kafatos F.C.;
RT　"Evolution of two major chorion multigene families as inferred from
RT　cloned cDNA and protein sequences.";
RL　Cell 18:1317-1332(1979).
DR　EMBL; K01841; AAA27779.1; -.
KW　Chorion; Multigene family.
FT　NON_TER　1　1
FT　NON_TER　33　33
SQ　SEQUENCE　33 AA;　2970 MW;　60D3208D2D16288B CRC64;

Query Match　　66.7%;　Score 116;　DB 5;　Length 33;
Best Local Similarity　78.6%;　Pred. No. 3.7e-07;
Matches　22;　Conservative　0;　Mismatches　0;　Indels　6;　Gaps　0;

Qy　　2 GLGYGGLGYGLGYGLGYGGLGYGGLGYGGLG 29
　　　　 | |||||||||||||||||||||
Db　　4 GCGCGLGYGGLGYGLGYGGLGYGSTG 31

RESULT　7
ID　P82170
AC　P82170;
DT　01-MAY-2000 (TrEMBLrel. 13, Created)
DT　01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT　01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE　CUTICLE PROTEIN 6.4 (LMNCP6.4).
OS　Locusta migratoria (Migratory locust).
OC　Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC　Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC　Acridomorpha; Acridoidea; Acrididae; Locusta.
OX　NCBI_TaxID=7004;
RN　[1]
RP　SEQUENCE.
RC　TISSUE=FIFTH INSTAR LARVAE CUTICLE;
RA　Anderson S.O.;
RT　Submitted (NOV-1999) to the SWISS-PROT data bank.
CC　-!- FUNCTION: COMPONENT OF THE CUTICLE OF MIGRATORY LOCUST WHICH
CC　　　CONTAINS MORE THAN 100 DIFFERENT STRUCTURAL PROTEINS.
KW　Structural protein; Cuticle.
SQ　SEQUENCE　52 AA;　6438 MW;　F3762E1E8F17D95B CRC64;

Query Match　　66.7%;　Score 116;　DB 5;　Length 62;
Best Local Similarity　63.9%;　Pred. No. 7.3e-07;
Matches　23;　Conservative　0;　Mismatches　0;　Indels　7;　Gaps　1;

Qy　　1 GGLGY------GGLGYGGLGYGGLGYGGGLGYGGLGY 30
　　　　 | ||| | |||||:| ||| ||||
Db　　21 GGYGYGYPGAYGYGYGGLGYRSLGYSGLGYSGLGY 56

RESULT　8
ID　P87483
AC　P87483;
DT　01-MAY-1997 (TrEMBLrel. 03, Created)
DT　01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT　01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE　KERATIN (FRAGMENT).
OS　Gallus gallus (Chicken).
OC　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC　Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC　Gallus.
OX　NCBI_TaxID=9031;
RN　[1]
RP　SEQUENCE FROM N.A.
RX　MEDLINE=85122780; PubMed=2578818;
RA　Wilton S.D., Crocker L.A., Rogers G.E.;
RT　"Isolation and characterisation of keratin mRNA from the scale
RT　epidermis of the embryonic chick.";
RL　Biochim. Biophys. Acta 824:201-208(1985).
DR　EMBL; M25642; AAA48931.1; -.
KW　Keratin.
FT　NON_TER　1　1
SQ　SEQUENCE　66 AA;　6813 MW;　E3C3A87D99918F25 CRC64;

Query Match　　62.9%;　Score 109.5;　DB 13;　Length 66;
Best Local Similarity　76.9%;　Pred. No. 4.1e-06;
Matches　30;　Conservative　0;　Mismatches　0;　Indels　9;　Gaps　6;

Qy　　1 GG--LGYGGL-GYGG--LGYGGL-GYGG--LGYGGL-GY 30
　　　　 || ||||| ||| |||| |||||| :||||| ||
Db　　1 GGSSLGYGGLYGYGGSSLGYGLYGGSSLGYGGLYGY 39

RESULT　9
ID　O76271
AC　O76271;
DT　01-NOV-1998 (TrEMBLrel. 08, Created)
DT　01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT　01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE　NONGRADIENT BYSSAL PRECURSOR.
OS　Mytilus edulis (Blue mussel).
OC　Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC　Mytiloidea; Mytilidae; Mytilus.
OX　NCBI_TaxID=6550;
RN　[1]
RP　SEQUENCE FROM N.A.
RX　MEDLINE=98393676; PubMed=9724735;
RA　Qin X.X., Waite J.H.;
RT　"A potential mediator of collagenous block copolymer gradients in
RT　mussel byssal threads.";
RL　Proc. Natl. Acad. Sci. U.S.A. 95:10517-10522(1998).
DR　EMBL; AF043944; AAC33847.1; -.
DR　INTERPRO; IPR000087; -.
DR　PFAM; PF01391; Collagen; 7.
SQ　SEQUENCE　904 AA;　77883 MW;　5529135651AD4C40 CRC64;

Query Match　　62.9%;　Score 109.5;　DB 5;　Length 904;
Best Local Similarity　82.8%;　Pred. No. 7.1e-05;
Matches　24;　Conservative　0;　Mismatches　4;　Indels　1;　Gaps　1;

Qy　　1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLA 29
　　　　 |||| ||:|||:|||| ||||| || ||||
Db　140 GGLG-GSLGGGGLGGGGLGGGLGGGLGGGGLG 167

RESULT 10
ID   Q9VKK5
AC   Q9VKK5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG13997 PROTEIN.
GN   CG13997.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dev I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003612; AAF52309.1; -.
DR   FlyBase; FBgn0031754; CG13997.
SQ   SEQUENCE 181 AA; 18540 MW; 4122603423B08FE8 CRC64;

Query Match   62.1%; Score 108; DB 5; Length 181;
Best Local Similarity 50.0%; Pred. No. 1.8e-05;
Matches   18; Conservative   6; Mismatches   6; Indels   0; Gaps   0;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGLGYGGLGYGGLGY 30
        | |:|:|| ::|| |:||:|:|| |:|:||| ::||
Db   72 GGPGFGPGFGPGFGPGFGPGFGPGFGPGFGPGFGPGF 101

RESULT 11
002049

ID   002049   PRELIMINARY;   PRT;   259 AA.
AC   002049;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   COSMID T20B6.
GN   T20B6.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstein L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Beck C., Wamsley P.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF000193; AAB52890.1; -.
SQ   SEQUENCE 259 AA; 23054 MW; C5F1690D2CE04CAF CRC64;

Query Match   61.2%; Score 106.5; DB 5; Length 259;
Best Local Similarity 68.8%; Pred. No. 3.9e-05;
Matches   22; Conservative   1; Mismatches   1; Indels   3; Gaps   1;

Qy   1 GGLGYGGLGYGGLGYGGLGYGGL---GYGGLG 29
        || |||| |||| | || |||| | |:|||:|
Db 110 GGGGYGGGGYGGGDGGGYGGYGGGYGGGYGGMG 141

RESULT 12
008631

ID   008631   PRELIMINARY;   PRT;   77 AA.
AC   008631;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   GLYCINE TYROSINE-RICH HAIR KERATIN PROTEIN.
GN   KRTAP6-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKIN;
RA   Aoki N.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D86419; BAA19680.1; -.
DR   MGD; MGI:1330228; Krtap6-1.
DR   InterPro; IPR001064; -.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.

KW   Keratin.
SQ   SEQUENCE   77 AA; 7742 MW; 99194ACFE2C34C29 CRC64;

Query Match   60.6%; Score 105.5; DB 11; Length 77;
Best Local Similarity   63.9%; Pred. No. 1.4e-05;
Matches   23; Conservative   0; Mismatches   6; Indels   6; Gaps   2;

Qy   1 GGLGYGGLGYGGLGYG---GLGYG--GLGYGGLG 29
          |||| |||||||| ||| |||| |||
Db   10 GGRGYGCCGYGLGYGYGGLGCGYGSYGCGYRGLG 45

RESULT 13
ID   O09048   PRELIMINARY;   PRT;   78 AA.
AC   O09048;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   GLYCINE TYROSINE-RICH HAIR KERATIN PROTEIN.
GN   KRTAP6-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKIN;
RA   Aoki N.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D86421; BAA19682.1; -.
DR   EMBL; D86420; BAA19681.1; -.
DR   MGD; MGI:1330228; Krtap6-1.
DR   INTERPRO; IPR001064; -.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW   Keratin.
SQ   SEQUENCE   78 AA; 8005 MW; 419194ACFF9C2FC CRC64;

Query Match   60.6%; Score 105.5; DB 11; Length 78;
Best Local Similarity   63.9%; Pred. No. 1.4e-05;
Matches   23; Conservative   0; Mismatches   5; Indels   7; Gaps   2;

Qy   1 GGLGYGGLGYGGLGYG---GLGYG--GLGYGGLG 29
          |||| |||||||| ||| |||| |||
Db   10 GGRGYGCCGYGLGYGYGGLGCCYGSYGCGYRGLG 45

RESULT 14
ID   O08636   PRELIMINARY;   PRT;   54 AA.
AC   O08636;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HIGH-GLYCINE TYROSINE KERATIN TYPE II.3 (FRAGMENT).
GN   KRTAP6-3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKIN;
RA   Aoki N.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D89901; BAA19687.1; -.
DR   MGD; MGI:1330279; Krtap6-3.
KW   Keratin.
FT   NON_TER   1   1
SQ   SEQUENCE   54 AA; 5729 MW; 2E7880E8B22C1787 CRC64;

Query Match   59.8%; Score 104; DB 11; Length 54;
Best Local Similarity   69.7%; Pred. No. 1.3e-05;
Matches   23; Conservative   0; Mismatches   4; Indels   6; Gaps   3;

Qy   2 GLGYGGLGYG-GLGYG---GLGYGGLGYGGLGY 30
          | ||||||| |||||| | ||| |||| ||
Db   9 GYGYGLGYGSYGCYGCGYGCGYG--GYGYGY 39

RESULT 15
ID   Q23052   PRELIMINARY;   PRT;   89 AA.
AC   Q23052;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   GLYCINE-RICH.
GN   T27E4.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Bradshaw H.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U64837; AAB04835.1; -.
SQ   SEQUENCE   89 AA; 9391 MW; A40C35A08A58494 CRC64;

Query Match   59.5%; Score 103.5; DB 5; Length 89;
Best Local Similarity   47.1%; Pred. No. 2.6e-05;
Matches   24; Conservative   0; Mismatches   6; Indels   21; Gaps   3;

Qy   1 GGLGYGGLGYGGL--------------------GYGGL-GYGGL-GYGGLGY 30
          ||| |||| ||||| |||| ||
Db   37 GGYGYGGYGGGCNCPPYAPCTHGPGGYGYGYGYGYGYGYGYGYGYGY 87

Search completed: April 24, 2001, 16:40:41
Job time: 431 sec